

Seppo Tiilikainen

Improving the National Cyber-security by Finding Vulnerable Industrial Control Systems from the Internet

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of
Science in Technology.

Espoo, 28.2.2014

Thesis supervisor:

Prof. Jukka Manner

Thesis advisor:

M.Sc. (Tech.) Timo Kiravuo



Aalto University
**School of Electrical
Engineering**

Author: Seppo Tiilikainen		
Title: Improving the National Cyber-security by Finding Vulnerable Industrial Control Systems from the Internet		
Date: 28.2.2014	Language: English	Number of pages:7+65
Department of Communications and Networking		
Professorship: Communications Networking		Code: S-38
Supervisor: Prof. Jukka Manner		
Advisor: M.Sc. (Tech.) Timo Kiravuo		
<p>Industrial control systems (ICS), which are used to control critical elements of the society's maintenance such as power generation and electricity distribution, are exposed to the Internet as a result of insecure design, and installation faults. Securing critical industrial systems is important for national cyber-security; malfunctioning elements in the critical infrastructure can quickly cascade into wide range of problems in the society. In the recent years increasing amount of cyber-attacks have been observed, and nations and criminals are developing offensive cyber-capabilities; industrial systems are also targeted as was seen with the Stuxnet-malware in 2010 causing havoc in an Iranian uranium enrichment facility.</p> <p>In this thesis a concept is presented to automatically find and evaluate exposed ICSs and report vulnerable devices to authorities for remediation. A prototype of the concept is built to prove the viability of the concept and to get data from port scanning real ICS devices in the Internet. With the prototype, 91 ICS devices were found out of the assigned 2913 IP addresses. Traffic volume produced by the scanner was insignificant compared to overall Finnish Internet traffic. The concept, called KATSE, is viable but not without challenges: ICS devices can definitely be identified from the Internet but analyzing the actual importance and purpose of the devices is difficult. Currently the Finnish legislation does not allow system intrusions or unauthorized security auditing even by authorities. Automated security auditing for the found devices would be useful to find the most vulnerable devices first but such auditing would require a change in legislation.</p>		
Keywords: cyber-security, industrial control systems, SCADA, fingerprinting, port scanning		

Tekijä: Seppo Tiilikainen		
Työn nimi: Kansallisen kyberturvallisuuden parantaminen etsimällä Internetistä haavoittuvia teollisuusautomaatiojärjestelmiä		
Päivämäärä: 28.2.2014	Kieli: Englanti	Sivumäärä:7+65
Tietoliikenne- ja tietoverkkotekniikan laitos		
Professuuri: Tietoverkkotekniikka		Koodi: S-38
Valvoja: Prof. Jukka Manner		
Ohjaaja: DI Timo Kiravuo		
<p>Teollisuusautomaatiojärjestelmiä, joita käytetään muun muassa voimantuotannon, sähkönjakelun ja jätevedenpuhdistuksen järjestelmissä, voidaan löytää julkisesta Internetistä. Tarve etähallinnalle ja keskittämiselle, sekä tuotteiden huono suunnittelu ja virheet järjestelmien käyttöönotossa, ovat altistaneet automaatiojärjestelmiä kenen tahansa ulottuville. Yhteiskunnalle tärkeiden kriittisen infrastruktuuriin kuuluvien järjestelmien turvalliseksi saattaminen on tärkeää kansalliselle kyberturvallisuudelle: ongelmat kriittisessä infrastruktuurissa voivat aiheuttaa voimakkaita häiriöitä eri puolilla yhteiskuntaa. Viime vuosina on havaittu kasvava määrä kyberhyökkäyksiä. Sekä rikolliset, että valtiolliset toimijat kehittävät kyberaseita ja myös teollisuusautomaatiojärjestelmiin on kohdistettu hyökkäyksiä. Vuonna 2010 Stuxnet haittaohjelma onnistui tunkeutumaan iranilaisen ydinpolttoaineenrikastamon järjestelmiin ja aiheuttamaan mittavaa fyysistä tuhoa.</p> <p>Tässä työssä esitellään konsepti, jonka avulla voidaan automaattisesti löytää haavoittuvia teollisuusautomaatiojärjestelmiä, ja raportoida löydökset viranomaisille jatkotoimenpiteitä varten. Työssä esitellään myös prototyyppi, jolla testattiin konseptin toimivuutta oikeilla suomalaisilla järjestelmillä Internetin yllä: sormenjälkitietokannan ja porttiskannauksen avulla 2913 IP-osoitteesta löydettiin 91 mahdollista teollisuusautomaatiolaitetta. Epäiltyjä teollisuusautomaatiojärjestelmiä pystytään löytämään Internetistä, mutta löydettyjen järjestelmien kriittisyyden ja tärkeyden arvionti ilman tunkeutumista kohteeseen on vaikeaa. Konseptia tehostaisi huomattavasti automaattinen tietoturva-auditointi, jolla tärkeimmät ja haavoittuvaisimmat kohteet voitaisiin paikallistaa ja poistaa näkyviltä nopeasti. Auditointi ilman järjestelmien omistajien lupaa vaatisi kuitenkin muutoksia lainsäädäntöön.</p>		
Avainsanat: kyberturvallisuus, teollisuusautomaatio, SCADA, sormenjälkeistys, porttiskannaus		

Preface

I would like to thank Timo Kiravuo for being an excellent instructor, and Jukka Manner for great advices and supervision of my thesis. I also want to thank Mikko Särelä and my other co-workers for their support throughout my time in the Comnet. Thanks to all my friends and family who have never doubted me and supported me during my hasty journey through studies. Especially I would like to thank my loving wife, Saija, for her tremendous ability to trust and to believe in me. Last, I thank with all my heart, my Lord and Savior, Jesus Christ, for everything.

Otaniemi, 5.2.2014

Seppo Tiilikainen

Contents

Abstract	ii
Abstract (in Finnish)	iii
Preface	iv
Contents	v
Abbreviations	vii
1 Introduction	1
2 Industrial automation in a modern networked society	4
2.1 Modern industrial control systems	4
2.2 Problems with externally networked automation systems	6
2.2.1 Exposure of industrial control system devices	7
2.3 Industrial control system incidents	8
2.3.1 Attacks towards industrial control systems	9
2.4 Nation-state cyber-capabilities	10
2.5 Summary	12
3 Finding specific devices from the vastness of the Internet	13
3.1 Identifying devices	13
3.2 Scanning the Internet	14
3.2.1 Analyzing the Internet Census 2012 data for serial port devices	16
3.2.2 Shodan, the search engine for Internet-connected devices . . .	17
3.3 Exposure of Finland	18
3.3.1 Results	19
3.3.2 Finland in comparison	21
3.3.3 Recap of the situation eight months later	21
3.4 Network traffic monitoring and warning system	23
3.5 Summary	24
4 Automated system to find exposed industrial control system de-	26
vices	
4.1 Overview of the concept	26
4.2 Scanning	29
4.3 The concept in detail	32
4.3.1 Overall architecture of the system	32
4.3.2 Components	32
4.3.3 ICS Device assessment	34
4.3.4 Practical issues of the concept	36
4.3.5 Geolocation	40
4.3.6 Additional functions	41
4.3.7 Possible problems in the concept	41

4.4	Legal issues	43
4.4.1	Finnish legislation related to communications	43
4.4.2	Legality of KATSE	45
4.5	Summary	46
5	Proof of concept	48
5.1	Prototype	48
5.2	Test environment	50
5.3	Testing	51
5.3.1	Scanning suspected industrial systems in Finland	52
5.3.2	Traffic and time	54
5.3.3	Analyzing port numbers	56
5.4	Traffic and time considerations on a national scale	57
5.5	Summary	59
6	Conclusions	60
	References	62

Abbreviations

ICS	Industrial Control System
PLC	Programmable Logic Controller
SCADA	Supervisory Control and Data Acquisition
RTU	Remote Terminal Unit
DCS	Distributed Control System
CPS	Cyber-physical System
CERT	Cyber-Emergency Response Team
NIST	National Institute for Standards and Technology (USA)
VPN	Virtual Private Network
HTTP	HyperText Transfer Protocol
FTP	File Transfer Protocol
SNMP	Simple Network Management Protocol
SSH	Secure Shell (Protocol)
IP	Internet Protocol
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
ICMP	Internet Control Message Protocol
ADDP	Advanced Device Discovery Protocol
NDU	National Defense University
OPC UA	OLE for Process Control User Agent
HTML	HyperText Markup Language
MAC	Media Access Control
ISP	Internet Service Provider
RIR	Regional Internet Registry
FICIX	Finnish Communications Exchange
DDoS	Distributed Denial of Service

1 Introduction

Today modern societies depend on a rich information infrastructure which is tightly linked to processes affecting the physical world, such as manufacturing, waste water treatment and power generation. Together, information systems, information infrastructure and real-world elements form the largest and the most complex structure built by man. It is a huge global cyber-space with a lot of players with different goals; both benevolent and malicious actors exist, just like in the Internet which can be seen as a sub-set of the global cyber-space. On national and international scale the cyber-space has a lot of potential to increase productiveness of the society but at the same time globally connected cyber-space provides means for malicious actors to support their own agenda, namely in the form of cyber-activism, crime, terrorism, espionage and high-profile nation state cyber-operations. The involvement of nation-state actors in cyber-attacks and espionage has raised the public awareness of the threats that the global cyber-space enables.

The Finnish cyber-strategy [1] has the vision of Finland being able to protect the vital functions of the nation in all situations against cyber-threats. According to the strategy, key factors in enabling that vision is having fluent coordination between all relevant actors and providing timely information about the state of the cyber-space and the attacks it is facing. In the national scale private sector companies, government agencies, educational entities and individual researchers are seen as relevant actors among which cooperation is required. The cyber-strategy defines situation awareness of the cyber-space as a key for detecting threats and countering them effectively. [1]

Motivation for this thesis rises from the need to protect the national cyber-space and especially the real-world elements linked to it. The cyber-strategy defines cyber-security as a state in which cyber-space is reliable and its operations are secured [1], and it acknowledges cyber-security as a vital part of the overall national security. Cross-relations between multiple parts of a functioning society stress the need to secure every element especially in the national critical infrastructure. Just like the whole society is dependent on electricity, the distribution of electricity depends on information systems. Attacking important information infrastructures can effectively disrupt electricity distribution which leads to more systems crashing and increasing dysfunction of the society. Industrial control systems are in the heart of the society controlling vital elements like power generation, distribution of electricity and water, waste water treatment and district heating. In this thesis a concept is proposed to enhance the national situational awareness of Finnish cyber-space by automatically finding vulnerable industrial automation devices and reporting their existence to authorities for possible remediation of the vulnerabilities.

Industrial control systems (ICS) as a term is used in this thesis to cover all control systems of any automated industrial installations including supervisory control and data acquisition systems (SCADA), automated manufacturing systems and different kind of process control systems using programmable logic controllers (PLCs) and remote terminal units (RTUs). ICSs can be local within one operational site or distributed (DCS) with multiple operating sites. SCADA systems are used to

monitor data coming from one or multiple sites and for controlling the automation systems in the sites. PLCs are equipment at the plant level controlling processes and physical equipment such as mechanical arms and valves. Along with PLCs, also RTU's are used in the plant level to gather data from sensors and transferring data to the SCADA system for automatic and human monitoring. All ICSs which affect the physical world through its functions are categorized as cyber-physical systems (CPSs). For example, an ICS system in a waste water treatment plant can consist of a SCADA system for monitoring and control purposes, and a number of PLCs which control physical equipment such as valves and pumps to regulate the water flow. [2, 3]

An example of an ICS is illustrated in figure 1. In the figure all of the components are an important part of the ICS. A process control network consists of an RTU reading data from the sensors, a PLC to control valves and pumps, and a local control station for human supervision and control. The SCADA server in the corporate network interacts with the process network, providing a supervisory control interface to the processes and allows collecting data for monitoring purposes. Database and historian servers also collect data from the process network. The three servers mentioned are located in the corporate network to make them easily accessible to office personnel handling the data. A firewall separates the process control network from the corporate network, and the corporate network from the Internet. Firewalls are protecting the networks from unwanted traffic, which is especially important for the delegate functions of the process control network.

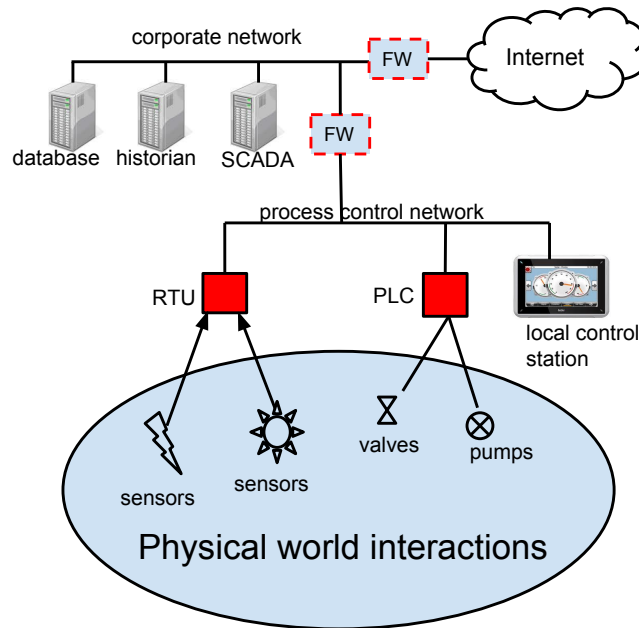


Figure 1: Illustration of an industrial control system

This thesis explores how an automated system for detecting vulnerable ICS devices from the Internet could be made, and would such a system be feasible con-

sidering traffic volumes, jurisdiction of authorities, privacy and legal concerns. The proposed system would help in securing assets with importance to the national critical infrastructure and previous research [4] and the cyber-strategy of Finland verifies that there is a need for enhancing the security of critical ICSs. A proof-of-concept prototype scanner of the proposed system was built to get validation for the concept and to get measurement data about traffic volumes and the speed of scanning a large number of IP addresses. The prototype was used to scan 2913 suspected ICS devices gathered from the previous research; the prototype was able to identify almost 100 devices of three different brands, based on pre-collected fingerprints. Observed traffic volumes were, as expected, really minor: approximately 17,3 kilobits per second during scanning. Even though ICS devices can be identified through fingerprints, their importance remains unclear as currently the Finnish legislation forbids any closer examination of systems without the permission of the owner of the target system. Some device enumeration can be done based on legally gained information such as IP address, domain name, owner of the domain, and so forth but analyzing the importance and functions of vulnerable ICS devices is definitely a research challenge for the future to make the concept more efficient.

In the next chapter, a closer look at industrial control systems and their current security issues is taken. Problems which cause ICS devices to be exposed in the Internet are covered, along with the possible consequences of malfunctioning ICSs. Also nation-state involvements in ICS attacks are briefly covered. Chapter 3 explains methods and tools used to find devices from the Internet. Previously researchers have used effective schemes to find ICS devices from the Internet such as the study which reported 7200 ICS devices part of critical infrastructure in the U.S. [5].

The fourth chapter proposes a concept to automatically and continuously find vulnerable control system devices inside a nation for the purpose of improving the national cyber-security. Vulnerable devices would be reported to their owners for remediation, decreasing the amount of easily targeted systems. The chapter views practical concerns such as time constraints for scanning, induced traffic volume, and legality of such a system. Chapter five explains a proof-of-concept prototype of the proposed concept. The prototype is tested against a privately hosted real-like virtual PLC device and against a set of real systems in the Internet. Finally, conclusions of the thesis and ideas for future work are presented in Chapter 6.

2 Industrial automation in a modern networked society

Critical functions of a modern, high-technology society are tightly linked together forming a large interdependent national critical infrastructure. Information infrastructure is at the heart of the society fusing formerly independent systems together: distribution of electricity is as dependent on communications networks as communication networks are dependent on electricity. Disturbances in some elements of the critical infrastructure can have massive ripple effects to other parts of the infrastructure, and towards the entire society. This chapter explains further the nature and importance of ICSs, and covers the reasons why control systems are being exposed to the Internet and what can happen when important ICSs malfunction. In the recent years an increasing trend of attacks towards ICSs has been discovered, including nation-state actors targeting specific industrial assets. Some reported attacks and the capabilities of nation-state actors are briefly covered to conclude this chapter.

2.1 Modern industrial control systems

ICSs range from manufacturing plants to critical infrastructure elements such as power plants, waste water treatment plants and dams, and the size and complexity of ICSs can vary greatly depending on the tasks of the operated site. Generally, the process control level of the ICS, where physical world interactions are made is very sensitive, and equipment is not directly linked to outside networks. As even a malformed request to a device in the process control can cause the system to halt [6], gateway devices and SCADA systems are used in between networks to gracefully allow monitoring and interactions with the delegate operations of the process control equipment.

This thesis focuses on the control and monitoring interfaces of industrial control systems, and to the exposure of those interfaces to outside threats coming from the Internet. The most typical interfaces to be exposed to the Internet are SCADA interfaces, PLCs and gateway devices between manufacturing and control networks. As the duty of an automation system is to be reliable and available at all times, confidentiality of the systems comes far behind [7]. ICSs have evolved from isolated systems into networked, often widely deployed, systems, and usually some ICS components, such as database servers and historians, are inside a corporate network. Even if an ICS is not directly connected to the Internet, a connection might still exist for example through a corporate network [2]. Figure 2 illustrates some interfaces of ICS components which might be vulnerable to outside threats. In the figure, the components of the process control network are the same as in Figure 1, except in the place of a firewall a gateway server or proxy device exists to allow remote controlling and monitoring. Process control networks usually use a fieldbus protocol such as Modbus for communications, and the gateway devices are sort of adapters making remote access, for example with TCP/IP, possible. If PLC, RTU and local control HMIs are TCP/IP-capable, they can also have remote management interfaces accessible with e.g. HTTP, Telnet or SSH protocols, exposing the devices to

the internet if not properly shielded.

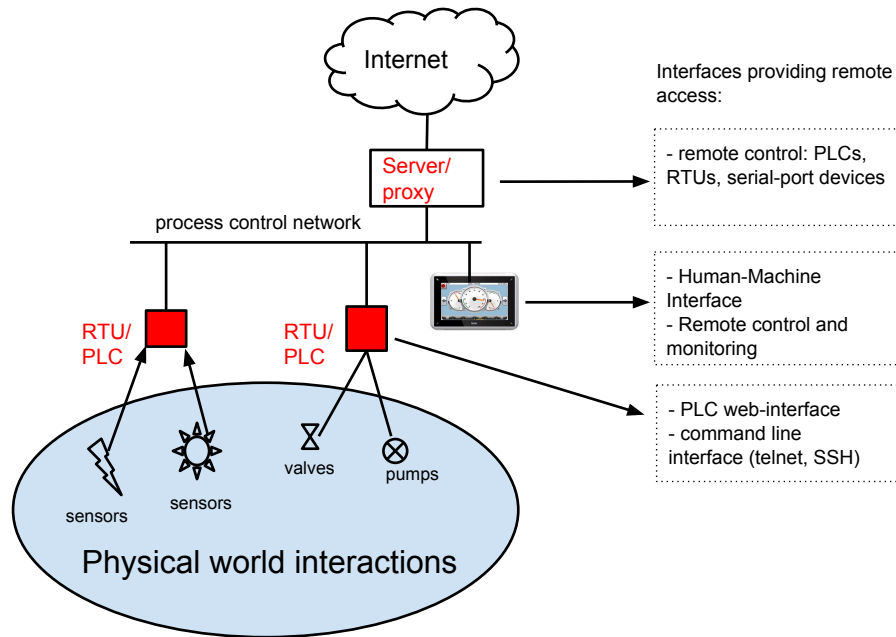


Figure 2: Possible exposed interfaces of ICS components

In the past ICS implementations have varied greatly, and systems have been specifically tailored for the target operating environment. Modern ICSs, on the other hand, are more and more similarly designed bulk systems using the same well-known manufacturers with the same device software. The need for inside information or detailed knowledge about the systems is increasingly not needed anymore to compromise ICSs. [8, 2] While this direction of development raises some security concerns, manufacturers do it for the profit. By selling ICS equipment off-the-shelf, vendors improve productivity and can sell the product with a lower price to the customers. Off-the-shelf products have bulk software and configuration without any customization, improving usability but leaving consideration about useful services and security measures fully to the customer. Major ICS vendors ship products all over the world, and they try to make the default configurations compliant for different kind of environments. The same flaws, default passwords and other vulnerabilities the equipment might have, are applicable all around the world. [2] Many ICS equipment ship out with a configuration of bad security settings, like default passwords and unnecessarily enabled services, making systems vulnerable instantly when they come into production, unless the devices have been reconfigured with a proper security insight. Some manufacturers have even included hard-coded backdoor management accounts into their devices without the option to disable the backdoor accounts. Although information about these backdoor accounts is not public, reports, have shown that secret accounts can be discovered, giving anyone with an access to the device a control over its functions. ICS Cyber Emergency Response Team (ICS-CERT) in the U.S. warned about medical devices having hard-coded

backdoor accounts [9], and researchers at security company Digital Bond found backdoor accounts in a widely used Schneider Modicon PLC controller [10].

2.2 Problems with externally networked automation systems

In the United States the National Institute for Standards and Technology (NIST) guide to industrial control system security [11] lists overall of 37 causes of vulnerabilities in ICS platforms, categorizing the vulnerabilities in configuration, hardware, software and malware related vulnerabilities. They also list 17 network related vulnerabilities, which largely derive from bad design, faulty configurations and overall lack of security knowledge. A big part of ICS vulnerabilities come from the mere fact that just like traditional IT systems before 1990's, ICSs in the past have not been designed from a security perspective, and a long life cycle of ICS equipment makes upgrading difficult. Availability, safety of personnel and equipment, fault tolerance from natural causes, and meeting regulatory requirements are primary concerns in industrial systems. Security has not been of major importance, even though data integrity and protecting intellectual property are considered important. The habit of not thinking about security affects the policies of companies: proper security management, auditing, documentation and training is lacking. Physical security of ICS assets are taken very seriously and that concern should also be extended to consider issues emerging from the cyber-space. [11]

Generally ICSs do not use encryption which makes them susceptible for traffic sniffing and forging commands. Access control is usually role based not user based, as it is more convenient to have one "always on" account for a monitoring station than each supervisor having an account of their own. In case of an emergency fast reaction to problems must be ensured and access control should not slow down the operators. The real-time, sensitive nature of many industrial systems makes patching software vulnerabilities difficult, leaving out-of-date systems like old Windows-based systems widely in the use. [12, 13]

Flaws resulting from bad security management and security principles, such as configuration errors, password policies and having unnecessary services enabled, can be fixed quite effectively with basic IT-security principles. However, some of the obvious flaws in ICS systems are not as easy to fix as they would be in traditional IT-systems. Automation systems in general, and especially process control systems, have special needs. Operations have strict time constraint, and real-time communications in milliseconds must be ensured even with equipment having low processing power, making for example cryptography and intrusion detection systems harder to implement without disturbing the system. High bandwidth is not required but otherwise high quality communications is expected: the systems generally do not tolerate jitter, packet loss or unexpected outages. [11]

Industrial plants have a high incentive to stay in production non-stop at all times. Even small down times in the production may result in significant financial losses. This makes patching and equipment upgrades very difficult for those systems, as they might have a maintenance down-time planned only once a year or once every two years. Replacing components with newer and more secure components

is not usually an option. The life-cycle of components in an industrial automation system can be multiple decades. The systems also consist of multiple interdependent components so replacing just one part of the system may not be possible. Because of specific equipment and maintenance contracts, a high customer lock-in exists and switching costs can be really high when loss of production and installation costs is considered. [8]

An informational source for remediating ICS security problems is the NIST guide to industrial control system security [11]. It offers valuable insight into the common problems in ICSs and into making a business case out of a good security portfolio. The guide explains several benefits for proper security management and discusses possible consequences for having unfixed vulnerabilities. Secure architectural designs are proposed with a defense-in-depth strategy having multiple restricted networks and a demilitarized zone (DMZ) governed by firewalls at the perimeter of each network.

2.2.1 Exposure of industrial control system devices

Evidence shows that during the past decade ICSs have been networked increasingly with systems outside of the control system [8]. Links between industrial and corporate networks are used to provide data for the office workers. Remote access to ICSs is made possible from anywhere through the public Internet or virtual private networks (VPN's). Widely distributed SCADA systems collect data from remote sites using cellular networks or traditional landlines. Exposing ICSs to outside threats can be really dangerous. This section covers some of the reasons why the devices are being exposed, unwillingly or by choice. [11]

Cost savings for centralization of control and monitoring can be substantial, especially for geographically scarcely distributed control systems. Instead of having a monitoring station and personnel on each site, SCADA systems can provide effective centralized management. Central monitoring and supervising also provides the ability to optimize the use of different assets. Furthermore, facts for decisions in a large scale are easier to visualize when all the data is readily at hand in one location. The benefits for networking control systems are understandable but arising security issues are not always understood.

Ignorance and lack of information security knowledge is a problem in the industrial systems engineering and management layers. As IT professionals with a good comprehension of security do not know enough of the nature of industrial systems, they are unable to design those systems. ICS engineers, on the other hand, have great knowledge of the systems, how they should run and what factors are important for personnel and equipment safety, but these engineers often lack the security knowledge. Generally the problem for an exposed ICS is on one of these three levels: design, implementation or management. Engineers designing the systems face the problems described above: knowledge from both automation system design and IT security would be needed to design safe and secure systems. Implementation of the devices into live systems faces problems of ignorance and belittling security threats. Engineers building the system and installing the components might have

clear and security wise solid instructions about networking of new devices. However, these instructions are sometimes discarded and proper security portfolio is not followed even if one exists [14]. In a project meeting about risk management in industrial control systems, an attendant (anonymous employee, personal communication, September, 2012) told a lively example of the gap between management personnel and engineers: he was observing an installation procedure at a remote site where the engineers installed a new device and then networked it to allow remote-maintenance and management at a later stage. When asked about the procedure of networking new devices, the engineers told that normally new devices should be cleared from the IT-support before installing devices into the corporate network but their experience was that IT-support would never give them the permission to network that specific device, so it was easier for everyone to just not tell anyone. [14]

In the vastness of the Internet of over three billion possible IPv4-addresses, one can question if networking random industrial automation devices is a problem. The Internet is indeed vast but so is the amount of tools and methods for exploring it. Search engines like Google and Bing can be used to find web pages containing certain information but they are not very efficient in finding specific devices, such as gateway proxies to industrial automation systems.

However, specialized search engines like Shodan [15] (www.shodanhq.com) and project ERIPP [16] (www.eripp.com) make finding all kinds of networked devices, which are not properly shielded against outside probing, very easy. Additionally, a lot of tools exist to discover and exploit industrial control system devices. Security group calling themselves "SCADA strangelove" [17] have released a tool to exploit devices using Siemens Simatic winCC platform, and also methods to discover industrial devices, such as Siemens S7 PLCs, using Modbus-protocol. They also released a brute force tool to crack the administrator password in the S7 PLCs. Open source security audit tool Metasploit also has modules for exploiting flaws in certain ICS devices.

2.3 Industrial control system incidents

The recent development to network-enable industrial control systems has led into highly increased amount of security incidents in the ICSs in the past 10 years. Companies who have suffered from security breaches do not want to report those incidents as it can be embarrassing for them and induce mistrust among their customers. Companies that do report incidents are punished for being insecure while companies denying everything are considered, often falsely, secure. The situation is problematic as statistical data and detailed information about cyber-attacks would help all companies in protecting against future threats. [8, 2] Companies and regulators must realize that the potential to cause accidents by attacking industrial control systems is real. Even though to date ICS attacks have not reported to cause any loss of lives, the potential is there. Targeted attacks may cause unexpected behavior and malfunctions similar way natural causes might induce accidents. In the past, malfunctions in SCADA systems have been responsible to cause huge damage and loss of lives. Some examples of major ICS incidents are presented below.

- In 1999 a failure in a SCADA system in Bellingham resulted in 230 000 gallons of gasoline being spilled, which then ignited, killing three people and injuring eight. [12]
- In 1992 a SCADA system monitoring a natural gas pipeline failed to recognize a malfunction in the system. Volatile liquid got ignited resulting in three deaths, 21 injuries and nine million U.S. dollars in damage. [12]
- In 2009 two metro trains in Washington D.C. collided, because a SCADA system failed to detect an idle train in the rail, resulting in nine deaths and 52 injuries. [12]
- A huge dam explosion in Russia 2009 killed 75 people and spilled 40 tons of oil into the river. Rebuilding the dam facility is still under construction and is estimated to cost overall of 1,2 billion U.S. dollars. The accident resulted from a turbine malfunction which made the automated safety system shut down allowing the accident to cascade. [18]
- The Tsernoby nuclear power plant explosion in 1986 is probably the most well-known ICS incident to date. Engineers ill-advisedly disabled the emergency cooling system so it would not affect the tests they were currently running. Obviously something else went wrong at the same time, causing the explosion which killed 56 people and is estimated to cause 4000 cancer related deaths while making the site uninhabitable for centuries. [19]

2.3.1 Attacks towards industrial control systems

Entities with different level of savviness are able to cause problems in ICSs. While state-level actors have the capabilities and resources to make serious well-planned attacks, also hobbyist, hactivists and criminals exploit the attack vectors of open ICS systems. According to a survey by Markey [20], companies part of the critical national infrastructure (CNI) in the USA are getting attacked constantly. Companies reported "daily", "constant" or "frequent" attempted cyber-attacks ranging from phishing attacks to malware infections and unfriendly network probes. One company reported being the target of approximately 10 000 attempted cyber-attacks each month. The categorization of what constitutes as a cyber-attacks may vary, but the survey proves that the companies are facing real threats and attacks on a regular basis.

Security researcher Kyle Wilhoit has been able to prove [21] that knowledgeable attackers are specifically targeting ICS devices. Using honeypots, i.e. fake systems that mimic real devices, Wilhoit was able to observe and record multiple targeted attacks to his honeypots posing as an ICS for a water plant. In a four month period he recorded overall of 74 targeted attacks of which 10 were knowledgeable enough to cease a complete control over the target system. Most of the attacks originated from Russia but half of the most sophisticated attacks came from China. In one particular incident, Wilhoit was able to trace the attacker to be the hacker group Comment Crew, also known as APT1. According to an intelligence report by security company

Mandiant [22], Comment Crew is likely a government-sponsored unit residing in Shanghai, China, using a large international infrastructure of computers in attacks against companies in various sectors in English-speaking countries.

In a case study of an incident in the Finnish Havaro-system designed to detect abnormal traffic in the networks of critical infrastructure companies, the sensor picked up malicious activity sourcing from the customer network. A host inside the network was infected with a malware, stealing 16 credentials used in operative systems, and 66 credentials used for personal services like Facebook and Gmail. [23]

Attacks can have sometimes really unpredictable results for unprepared organizations. A U.S. company The Economic Development Administration (EDA) which is a part of Department of Commerce, got informed from the Department of Homeland Security (DHS) that they had a malware infection in their systems. They responded by cutting off their systems from the Internet. After considerations, the CIO fearing from being attacked by a nation-state actor, decided that their computers had to be physically destroyed to make sure the malware was gone for good. This resulted in computers, printers, keyboards and mice worth of 170 000 U.S. dollars being destroyed and a full recovery of installing new systems and security auditions took nearly a year. Later audit revealed that the detected malware was a common non-targeted malware. [24]

2.4 Nation-state cyber-capabilities

The great potential of cyber-space is also realized by state-level actors like governments and well-funded military and intelligence organizations. Many nations are adding cyber-space as part of their military doctrine [25] and the development of both defensive and offensive cyber-capabilities are being funded by many countries [26]. Five major players, USA, China, Russia, Israel and France, are leading the way of building advanced offensive cyber-weapons. [19] As of 2006, USA has recognized the cyber-space in its strategic doctrine as equally important as land, air, and sea space, and have since established a dedicated strategic command for cyber-space operations. The UK also has a center for security operations and an office of cyber-security in the cabinet office. [27] South Korea and India have lately increased their effort to prepare to deal with issues in the cyber-space, and Russia and especially Israel are known to possess strong cyber-capabilities. [25]

From the year 2000, state-sponsored attacks have been reported to disrupt or disable opponents' systems. In 2000 Israel defaced public Hezbollah and Palestinian websites, resulting in counter attacks disrupting Israeli financial institutions and government computer systems. In 2001 China, following a maritime dispute with the U.S., allegedly attacked a Californian power plant causing disturbances in the electricity grid. Israel used a cyber-attack against the air defense network of Syria to assist in bombing a nuclear plant construction site in 2007. Similar exercises of state-level cyber-capabilities have since been increasingly observed. While cyber-attacks by a nation-state can be considered as an act of war, espionage in cyber-space, like traditional espionage, is allowed by international treaties (Tallinn-manual, laws of armed conflict [28]). However, attribution of attacks can be really difficult, and

actors in most cases can deny participation. [25]

Tracing the attacks in the cyber-space is a lot more complicated than in regular real world attacks, which makes the line between espionage and attacks a tempting one to cross. While attacks and espionage campaigns can be traced somewhere, plausible deniability always exists in cyber-space, as origins of activity can be spoofed, proxied, routed, concealed and so on. One example of espionage crossing the line of attacking is from the U.S., where the electricity grid was penetrated by spies; the attack seemed to originate from China and Russia, along with some other countries. The spies penetrated the systems and left behind software capable of disrupting the operations of the systems. Security experts evaluated that the purpose of the attack was to map the critical infrastructure elements in the U.S. and the sophistication level of the attacks points to China or Russia, although no nation-state involvement can be confirmed. [29]

Cyber-space is also realized to be an effective part of modern warfare, as is discussed in a journal paper by Deibert [30]. In an armed military conflict between Russian and Georgia in 2008, cyberspace played an important role and both sides used it to their advantage, for example by using the Internet, TV and other information channels to spread influential information, news and rumors to sympathize their cause. Russia did this to the extent that Georgia decided to sensor Russian TV channels and prevented access to some Russian web sites. Both sides, although government involvement can be denied, attacked information infrastructures and government computer systems of the opposite side. Large scale distributed denial-of-service (DDoS) attack, and malicious hacking, originating from within Russia, put down several Georgian web sites. Actors behind the cyber-offenses might be civilians acting on their own but the scale of the attack suggest that Russia encouraged the attacks or did not want to prevent the attacks. The DDoS attacks were delivered by a huge botnet consisting of millions of compromised computers. [30]

One of the most sophisticated and well-known cyber-attack was the Stuxnet worm disrupting Iranian uranium enrichment plant in Natanz. Stuxnet was the first complex malware targeting specifically industrial control systems with many sophisticated methods like zero-day exploits, Windows and PLC rootkits, and anti-virus evasion techniques. The initial spreading method of Stuxnet was from infected USB-drives, making Natanz facility reachable even though it was not connected to any external computer network. The infected USB-drives exploited a Windows automatic file execution vulnerability to immediately spread to computers the USB-drive was connected to. Once inside a computer, Stuxnet spread inside the local area network trying to find systems matching its objectives. The primary goal of Stuxnet was to infect specific Siemens control system equipment, and to reprogram PLCs responsible for the operations of the ICS. The creators of Stuxnet, U.S and Israel, wanted to be able to covertly infect and reprogram the PLCs of the Natanz enrichment plant without being detected. In a tactical sense Stuxnet was successful: approximately 1000 of the total of 5000 centrifuges were destroyed after the infected PLCs made them spin out of control without the operating staff knowing malfunctions were happening. However, long term implications of the malware can only be guessed. After U.S. and Israeli involvement in Stuxnet was confirmed by

unnamed officials who spoke to New York Times [31], Stuxnet has started a process of proliferation of the usage of cyber-weapons, while further endangering the relations between U.S. and Iran. [32, 26]

2.5 Summary

Tightly linked infrastructure elements of the society along with the trend to network-enable industrial control systems has led to a situation where possibly crucial control systems are exposed to the public Internet and are visible to anyone. As manufacturers seek financial efficiency by designing bulk systems with defective security configurations, the burden of security is solely on the often security-ignorant engineering personnel implementing new devices into production environments. Even if a proper security portfolio exists to maintain good security practices, new devices get networked as the result of misconfiguration or belittling of threats by personnel doing the implementation. Exposed ICSs do pose a threat as can be seen from the hazardous accidents malfunctioning ICSs have caused. Adding the increased trend of attacks towards industrial systems, and the fact that tools and methods exist to find specific targets from the Internet, as explained in the next chapter, threats towards industrial control systems are real.

3 Finding specific devices from the vastness of the Internet

Internet has evolved from a network of personal computers, routers and servers into a so called Internet of things [33] with billions of devices: smart phones, smart meters, web cameras, televisions and so on, all connected to the fabric that is the modern Internet. Even in a national scale, the amount of networks and Internet-connected devices are measured in millions, and finding a specific device without knowing the IP address of the device is like trying to find a needle in a haystack. However, a lot of tools and techniques have been developed to aid in finding, identifying and securing devices in the vastness of the Internet. Plethora of scanning software exists to map networks and find open ports in devices for communications. When considering national security in the cyber-space, the ability to form a comprehensive situational awareness of the state of the nation's cyber-space can be crucially important, albeit very difficult. Important cyber-space elements, such as national critical infrastructure is operated by multiple organizations, individuals, governmental and private entities. Overseeing everything to form a collaborated sense of situational awareness is close to impossible: motivations, jurisdictions, resources and unwillingness of disclosures are notable obstacles.

3.1 Identifying devices

In this thesis the term "scan" or "port scan" is used to describe an action where a target computer is being sent requests in order to learn information of the target computer, such as open ports and used services. A port scan of a target host might reveal that, for example, an SSH service is online and responsive on port number 22 and an HTTP service is responsive on port number 80, and that no other queried services are responsive on the host. Internet Security Glossary (RFC 2828 [34]) describes port scanning as "*an attack that sends client requests to a range of server port addresses on a host, with the goal of finding an active port and exploiting a known vulnerability of that service*", but in many cases port scanning is not used for attacking but for inspection to improve security. In this thesis port scanning is seen as a neutral method of gathering information rather than an offensive action on default. If the intention of a port scan is to exploit the target system, it can be classified as an attack and can be held against the court of law.

Scanning IP addresses and finding open, responsive ports is easy. The difficulty in having a useful scanner is the identification of devices behind the IP address that is queried. Through communications, an absolute confirmation which kind of physical device is responding can never be made with 100% accuracy, as the software installed in the device does the communications tasks: outputs and responses of the software can always be forged and made to pretend to be something that it is not. However, in many cases, observing the responses of the software a probable identification can be done about the purpose of the device. The process of gathering these software responses, and then comparing them to previously known responses for identification, is called fingerprinting.

Fingerprinting can be done multiple ways, but the goal is to have a unique set of fingerprints so that they can be used in identifying similar devices. Popular open source port-scanner software Nmap (www.nmap.org) has a device fingerprint database of over 2000 entries. Nmap uses the TCP/IP-stack implementation of software in fingerprinting. It is accurate in identifying devices in the fingerprint database, but when dealing with publicly less known devices, like most ICS devices, their TCP/IP-stack is unknown to Nmap and hence it cannot be used in identification. A higher-level approach for fingerprinting is simply looking at the banners and headers sent by the device when querying services like HTTP. For example, most web-servers reveal in the HTTP-header the server software and software version. The same goes for ICS devices with web-servers embedded on them. The response message can tell that the device is, for example, a Schneider SCADA-server version 1.0.

Wisely constructed HTTP-headers do not give away too much information, resulting in too vague responses to draw any conclusions from them. Alternate methods still exist to identify devices. A research to find ICS devices in Finland [4] suggests that even though some devices might have general HTTP- and FTP-responses, the devices might still give away their name and model with other services, mainly with netBIOS or SNMP messages. Both netBIOS and SNMP are protocols used inside a private, trusted network, and should not be visible to outside queries. However, erroneous configuration at the local network devices can extend the protocol advertisements outside the private network, making the devices detectable by outsiders. NetBIOS protocol often advertises the unique MAC-address of the device, which can be resolved to reveal the manufacturer of the device.

A concept proposed in the next chapter describes how fingerprints from multiple sources along with location and owner information from an IP address can work in conjunction to help identifying ICS devices and assessing their use environment when randomly scanning the IP address space of a nation.

3.2 Scanning the Internet

Scanning the Internet of over 3,6 billion possible IP addresses in the past has been quite a time consuming tasks. In 2010 Electronic Frontier Foundation (EFF) conducted an Internet-wide scan to gather data on the use of encryption online. They used the popular network scanner Nmap and their scan of the Internet took two to three months with multiple Nmap instances. Recently, a new line of scanners have been developed to enable scanning large networks in a reasonable time. Given a high-speed gigabit source network, ZMap [35] and Masscan [36] scanners are able to scan every host in the entire Internet in less than one hour. Clearly, the vastness of the Internet is no longer providing security through obscurity the way it used to before efficient port scanners. Next in this chapter, some Internet-wide studies done by scanning are presented. Researchers have used different kind of methods but all of them have been able to enumerate devices connected to the Internet quite efficiently.

One of the most extensive public research papers on Internet-connected devices is

the project called Internet Census 2012 [37], where an anonymous researcher scanned and recorded information from every Internet Protocol version 4 (IPv4) address in the world, during a period of nine months. He queried roughly 3,6 billion IP addresses multiple times and presented all gathered information for free to download. The records consists of HTTP-headers, ICMP responses, reverse DNS records, trace route information and service probe information from Nmap scanning software. The terabytes of gathered data is available at the web site of the project [37].

The researcher started his project by scanning random IP addresses for open telnet ports, and trying out a few different combinations of default credentials, such as "root/root" and "admin/admin", to try to log into the found hosts. He then, having administrative privileges on the target machine, uploaded a simple script which started to scan for more open telnet ports in random IP addresses. Using this methodology, he exponentially increased his scanning capability until a massive distributed scanner network of over 400 000 devices was at his command, which the author named as Carna botnet. With all of the hijacked devices doing the distributed scanning tasks, every possible IP address was queried within one hour.

The compromised devices were found from all over the world, and they were mainly Linux-based embedded devices, such as digi boxes, routers of different scale, and ICS devices. Although the compromised hosts allowed login with default credentials and the uploaded scanning software was intended to do as little harm as possible, the actions of the researcher to hijack the hosts would be considered illegal in most countries, which is probably why the author decided to publish the research anonymously. To prevent harm or disturbance to the hosts or private networks behind them, he used the hosts only for external scanning and programmed the scripts to delete in case of a device reboot.

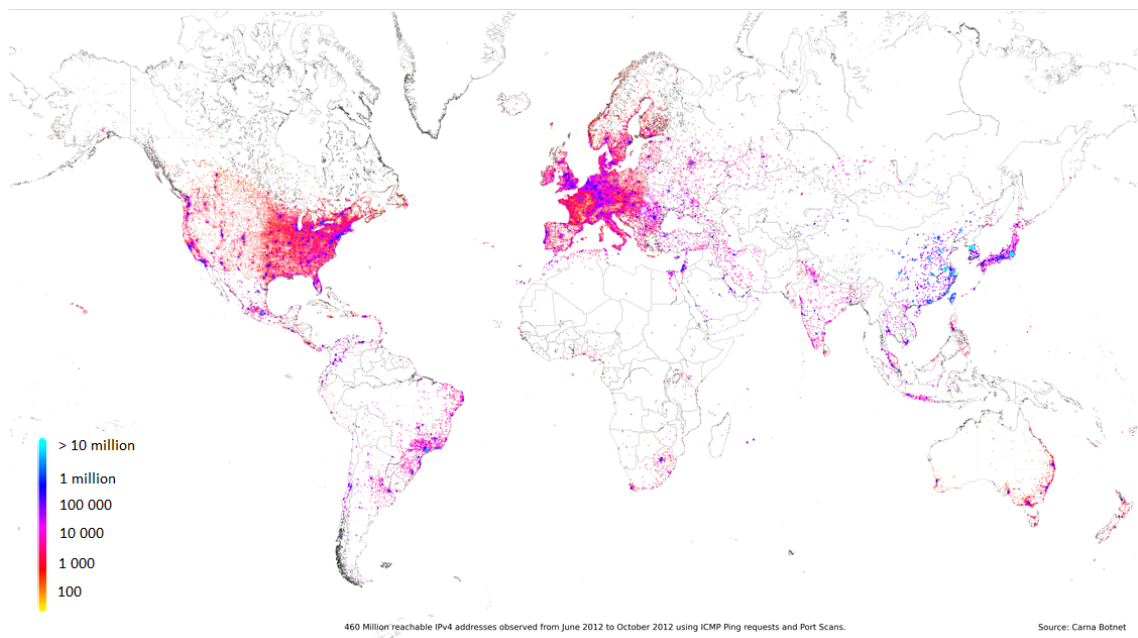


Figure 3: A world map of IP addresses which answered to ICMP-pinging [37]

The research resulted in terabytes of data about the usage of IP addresses, software, ports and services. According to the paper approximately 1,3 billion IP addresses showed signs of use while the rest 2,3 billion did not. Out of those 1,3 billion addresses, 460 million responded to an ICMP-ping: a global distribution of responsive hosts can be seen in Figure 3. The collected records provide valuable information for researchers. For example, using the data set from the Internet Census 2012 project, security researches at Rapid7 were able to find 114 000 serial port servers used in ICS networks, and were able to use the data to construct modules to scan for such devices. As the Census data include used port numbers for services, the data can be used to find out the ports on which e.g. ICS device remote access services are being run on. That can be especially helpful in finding devices which run services on non-default ports. The data set can also be used to identify fingerprints of the devices based on the services they run, which would be valuable information for systems trying to identify vulnerable devices, like the KATSE-concept proposed in the next chapter.

3.2.1 Analyzing the Internet Census 2012 data for serial port devices

Researchers from a security company Rapid7 [38] used the data-set from Internet Census 2012 to identify 114 000 serial-port servers manufactured by Digi International and Lantronix, two major companies in providing third-party solutions in connecting serial-devices into TCP/IP networks for example for remote management purposes. So called serial-devices are used in many kind of systems; industrial control systems use them for connecting older equipment lacking TCP-capabilities to larger networks. For example sensors, remote terminal units and PLCs might have only serial-ports to connect them into the control network. Serial-port servers are used as adapters to provide network connectivity, remote management and maintenance to the serial-devices. Usually serial-port servers accept outside connections through telnet, SSH or a web-interface, allowing control of the device after authentication. In some cases when the serial-devices have to be reached with a direct link, serial-port servers can be setup as proxies, directing connections to the correct port directly to the serial-device. Some vendors require specific software for remote user to be able to communicate with the serial-device. Vendors use proprietary protocols, which increase the security of the remote access as long as it is unclear how it communicates. Digi uses a proprietary protocol RealPort in their serial-port servers, and the security researchers from Rapid7 have been able to identify them from the Internet Census data-set.

A big problem with serial-port devices is that they trust any device that is connected to them. This means that a directly connected serial-port server is always trusted, and getting an access to the often weakly secured serial-port server provides automatically access also to the serial-port device behind the server.

The researchers were able to find Digi and Lantronix devices from public SNMP announcements, in the Internet Census data-set. For example a Digi serial-port server might advertise its full name in the SNMP announcement, making them easy to identify. 114 000 unique IP addresses belonging to Digi and Lantronix devices

were found, of which approximately 95 000 were used over a mobile network connection, such as GSM or 3G. In addition 14 000 unique IP addresses were found which ran a proprietary protocol called Advanced Device Discovery Protocol (ADDP), developed by Digi. ADDP is a protocol with known exploits and it is not usually needed by the system. ADDP only makes identifying and exploiting Digi-based serial-port devices easier. In addition to above mentioned, FTP banners proved useful, providing identification of another 8000 Digi devices. Overall, 13 000 devices allowing access without authentication were identified. Results indicate that quite a few Digi and Lantronix devices are being run on ports 2001-2010 and 3001-3010, and scanning those ports should be useful when trying to find vulnerable ICS devices.

3.2.2 Shodan, the search engine for Internet-connected devices

The Shodan search engine [15] scans the entire Internet in a randomized fashion and keeps a database of everything it finds. To be able to find different kind of devices Shodan port scans every possible IPv4 address for specific well known ports which are used by popular services, such as HTTP and SSH. As an example, Shodan finds responsive Telnet-services by querying the port 23 on the target host with Telnet protocol. If the host has no authentication requirements, Shodan automatically establishes a connection to the host. Shodan does the same thing for FTP-connections on port 21. If the FTP-server allows anonymous log in, Shodan logs in and records available commands in the system. According to the creator of Shodan, John Matherly [39], Shodan uses multiple distributed servers for scanning, and is able to scan the entire Internet once a month.

Shodan started off in 2009 by querying just four different ports: 21 for FTP, 22 for SSH, 23 for Telnet and 80 for HTTP. Since that, Shodan has grown considerably, now scanning 33 ports from each host it finds [39]. The most popular ports that Shodan scans can be seen from Figure 4, along with the respective number of hits for each port world-wide.

Information from Shodan can benefit a lot of people with different kind of interests. Security researchers get important empirical data about what kind of devices are online in the Internet. Discovered vulnerabilities in software is one thing but actually getting data about how widely that vulnerable software is used is very important. Security managers can use Shodan to check for their organizations domain for any unwanted exposed devices. Software companies can get valuable data of how much and where their software is used. Naturally also malicious actors can benefit from Shodan but the security awareness Shodan is helping to raise, hopefully outweighs the bad things it enables. Without Shodan, criminals and malicious actors would still find vulnerable targets with other means and exploit them. Shodan provides valuable empirical data about the exposure of devices and helps researchers prove that e.g. in the case of exposed ICSs, something has to be done.

Security researchers in different fields have found Shodan very useful because it has a constantly updating and expanding database about Internet connected devices. In 2011 Eireann Leverett from Cambridge University presented in his thesis [14] the findings of several thousands of exposed industrial control systems all over the world.

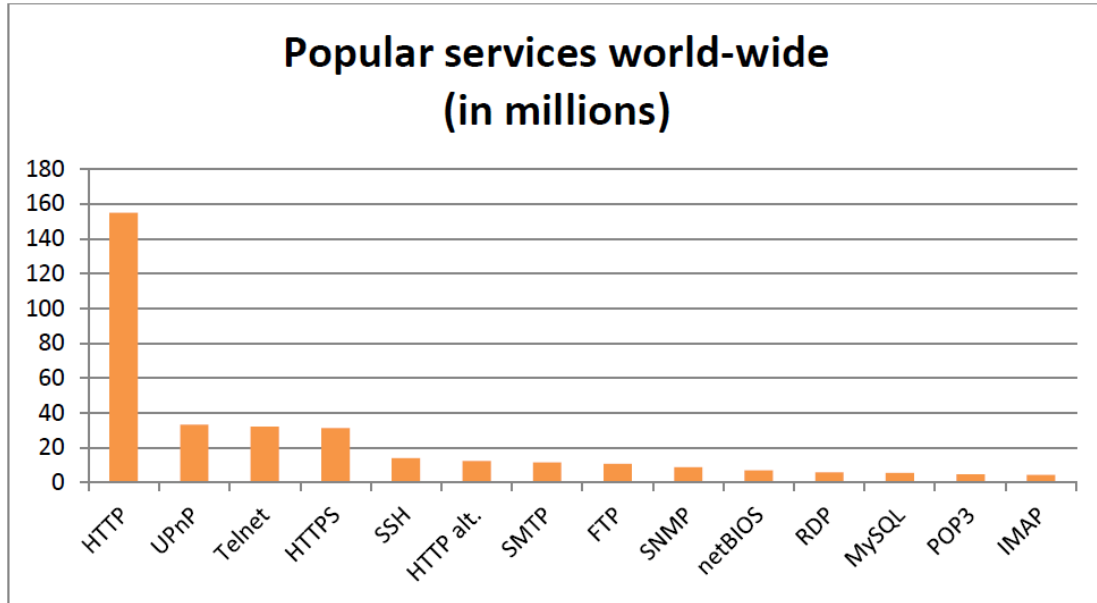


Figure 4: Popular services world-wide from the Shodan search engine (Oct. 2013)

Leverett used Shodan to find the devices and to map exposure, vulnerabilities and location of the ICS devices. From previous experience with control systems he derived a search word list of 27 items related to devices used in ICSs. He mapped the results to a visual representation in a world map, with a time reference, showing the location and vulnerabilities of each found device sorted by a date they were found by Shodan. Before submitting his thesis, Leverett reported his findings to the ICS-CERT authority in the U.S.

Project SHINE (Shodan Intelligence Extraction) [5], a research carried out with the help of the Department of Homeland Security, identified a huge number of ICS devices using Shodan. Searching globally, with a much larger search word list than what Leverett used, they were able to find 500 000 IP addresses suspected to belong to ICS devices all over the world. With the help of ICS-CERT, they were able to determine that at least 7200 of the found addresses were directly related to devices that are part of critical infrastructures in the United States alone. Although they admit recognizing devices being part of the critical infrastructure is hard, they claimed a high probability that most of the 7200 devices were actually important ICS devices used in critical infrastructures and should not be publicly reachable. After the critical devices were identified, ICS-CERT warned authorities in more than 100 countries about their Internet-connected industrial control systems.

3.3 Exposure of Finland

In the fall of 2012 a research project in Aalto University [4] began to map the exposure of Finnish industrial automation systems. Studies by others suggested that vulnerable ICS devices exists everywhere in the world and that search engines like Shodan and ERIPP made them possible to locate. The research was conducted

with Shodan, focusing on industrial automation systems, power management systems, building management systems and remote access servers for industrial control systems. With customized search queries on the Shodan nearly 3000 devices of interest were found. A check-up of the results revealed that 1968 of the devices were still responsive at the time of the research. A report on the research was published in the March of 2013. An anonymized version of the report was publicly distributed and authorities were given a version with a detailed list of IP addresses of every found device for possible remediation. Also, in quarter four of 2013, the search was performed again with the same keywords to observe the change in results.

3.3.1 Results

Using 41 distinctive search queries, Shodan found 2915 unique IP addresses belonging to suspected ICS devices used in Finland, categorized in Table 1. A majority of them, approximately 2300 devices, belonged to building automation and power management systems, and the rest were devices used in industrial and commercial automation systems, from manufacturing and district heating to supermarkets and retail businesses. By comparison to other countries such as Germany, France and United States, Finland was found to have quite a lot exposed automation systems: per capita statistics for exposed devices were 0,6 devices for Finland and below 0,2 for the other mentioned countries. Some explanations might be the low population density of Finland along with great communications infrastructure to support networking of ICS devices. On the other hand if the know-how to network ICS devices exists, proper security knowledge is either ignored or not present in the implementation of new networked ICSs. During the writing of this thesis a check-up after approximately eight months later than the original research was made, revealing that not many of the systems were taken offline and in addition, Shodan was able to find more devices; an increase of roughly 60% in results was observed.

Table 1: ICS devices found in Finland with Shodan

	Q1	Q2
Total	2915	4695
SCADA/ICS	77	95
Industrial	540	574
BMS	2298	4026

Search queries were customized by hand, going through popular automation equipment manufacturers and manuals to find information useful in fingerprinting the device. Only queries resulting matches in Shodan were kept: hundreds of queries which did not yield any matches, were tried during the research. 77 devices related to SCADA and industrial-grade automation devices were found. Those included different kinds of communications and control and monitoring equipment used in industrial systems. Not only were the devices exposed, some of them had no access control to restrict unauthorized access. Eight command line control interfaces for

automation systems were found over a Telnet-access on port number 23, and none of those required any authentication to access them.

540 pieces of equipment used in industrial automation, retail businesses such as supermarkets, shopping malls, and building automation were discovered. Most of these devices are adapters to enable communications of old devices with Ethernet networks. These adapters make the devices able to communicate in modern TCP/IP networks and allow remote controlling and remote data gathering. Found devices belonged to a variety of organizations, mostly big retail companies but also some energy and water treatment companies were identified.

Rest of the found devices were from building management systems (BMS) and building-scale power management systems, accounting for almost 2300 devices. BMSs are used to control automation components in buildings including physical access control with electronic door locks, heating, venting and air condition systems, lighting and alarm systems. Most of the found systems belonged to apartment buildings, stores, offices and gas stations. Also some peculiar devices were found belonging to an ice rink, a prison, a hospital and a bank office. Having the BMS of a hospital exposed to the Internet is especially worrying, since hospitals generally have very strict heating and air condition requirements and tampering with those controls could put people's lives at risk. Recently a report of a backdoor account vulnerability in a building management system called Niagara Framework, from Tridium, has got a lot of attention in the USA [40]. The Niagara Framework is widely used in hospitals and in government and army buildings. The vulnerability allows unauthorized access to the system via the backdoor account, i.e. hard-coded credentials, which cannot be disabled without patching the software. In Finland, 252 devices using Niagara software were located, of which at least 184 used the older software version that has the backdoor account vulnerability. The found power management devices were building-scale power distribution units (PDU), and UPS-control devices. These devices are used for example in data centers. Found devices belonged mainly to hotels and hosting companies.

One of the most alarming devices found was a remote control interface for an automation system controlling a critical element of railway traffic. Without further analysis the potential of this control interface was unclear, but surely it should not be accessible from the public Internet. Below are some anonymous examples of the environments where some devices were found.

- An open telnet-port in a device with references to a windmill
- A device in a small power plant responds to netBIOS queries. May allow sniffing of internal IP and MAC addresses and credentials.
- A router-firewall device of a water treatment company had stored credentials in the web management interface
- An open telnet-port in a building management system in a prison
- Exposed web-based control interface for a building management system in a bank office

3.3.2 Finland in comparison

To find out how exposed ICSs are in Finland in comparison, we chose eight countries which share the level of industrialization with Finland, along with good percentage of Internet coverage. Shodan was used to find ICS devices based on our own search keyword list: overall 53 search terms were used, 33 from our list and 20 from the research of Leverett to reduce the bias of having a list of too much devices used only in Finland or in the Scandinavian market. Overall 132 775 IP addresses were found globally. As the first chart in Figure 5 shows, majority of them, 41,1%, are located in the USA. Percentage-wise Finland is second to last among compared countries only France having less exposed devices. On the other hand if we look at the per capita statistics Finland has the most devices found per 1000 inhabitants, as the second chart illustrates, by quite a large margin. Some bias towards Finland and Sweden is possible due to specialized search terms but it does not explain such huge differences. Even if we drop out building automation devices, Finland and Sweden, with equal per capita percentage, have twice the amount of exposed ICSs compared to the USA.

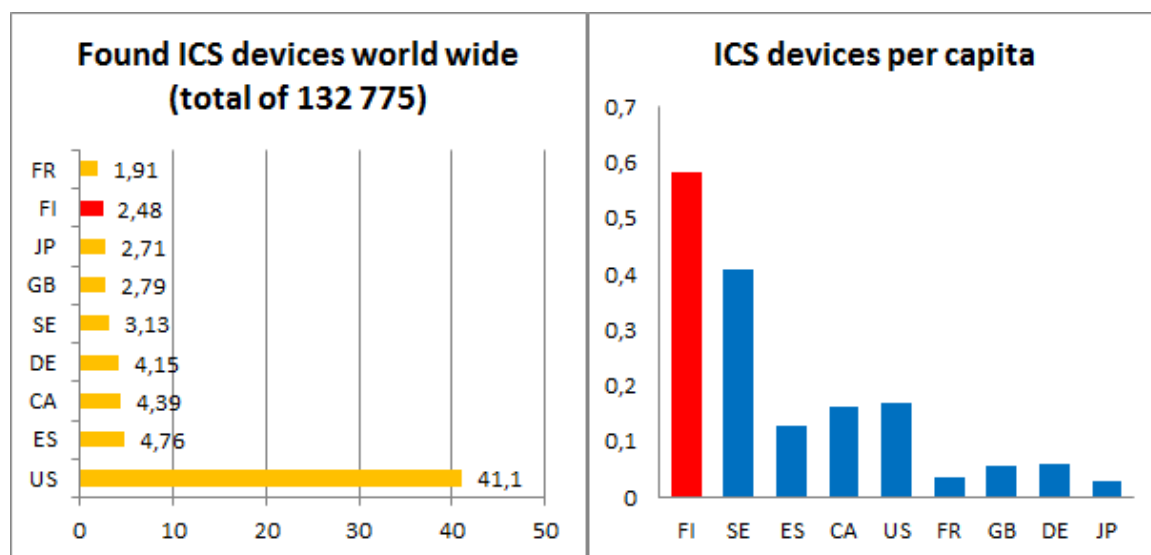


Figure 5: First chart: Percentage of all found devices for countries. Second chart: Found ICS devices per 1000 inhabitants

3.3.3 Recap of the situation eight months later

The original research was done during the first quarter of 2013 (Q1/2013) when 2915 unique IP addresses were identified with Shodan, of which 1968 were still found to be responsive (i.e. online) during the writing of the report. The responsiveness of the hosts was done with the Nmap network scanner. After eight months (Q4/2013), the same hosts as before were scanned again to see if any improvement had happened. Scanning the same 1968 hosts as in Q1, 1602 of them were still online, making 366 more offline hosts than before. However, when scanning the original IP address set of

2915 addresses, a two run average of online hosts was 1858. Decisive conclusions are hard to make, since some of the IP addresses might have been dynamically allocated and since changed; a different device might currently be holding that specific IP address instead of the device identified with Shodan. Some devices might have been protected against scans or taken offline on purpose but to get a conclusive analysis why some hosts are no longer online, a further analysis of comparing the results between Q1 and Q4 should be made.

In Q4, Shodan was queried again with the same set of keywords as before in Q1. A large 61% increase in found devices occurred, as can be seen from Figure 6 below. Results are divided into two categories: building automation and industrial automation. While only a minor increase was seen in industrial automation, building automation expanded greatly causing most of the 61% increase in the overall results.

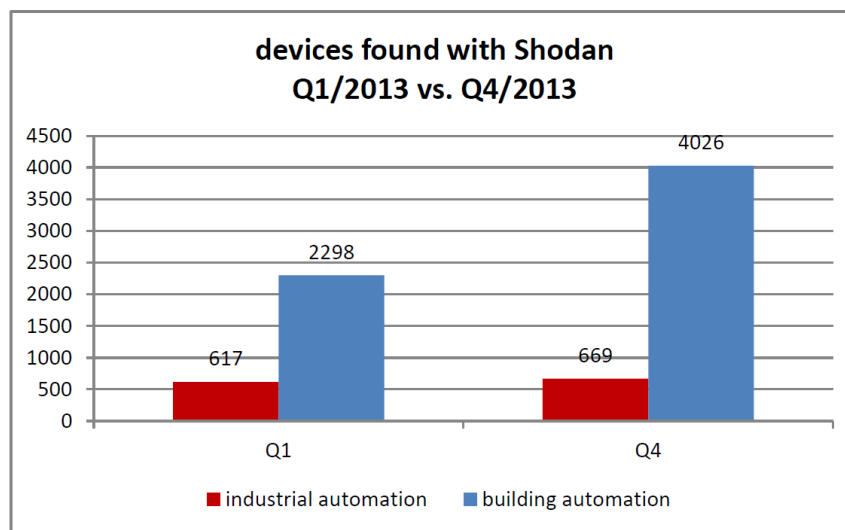


Figure 6: Comparing devices found with Shodan on Q1/2013 and Q4/2013

In Q4 of 2013, 4695 devices were identified with Shodan. A similar responsiveness check-up as was done in Q1 was performed for the new results; 3281 devices were found to be online. In the report in Q1 of 2013 Shodan was estimated to have scanned somewhere between 20-30% of the entire IPv4-address space of Finland. In hindsight Shodan was at that time heavily speeding up the speed of its scanning efforts, probably having scanned over 50% of Finland's IP addresses already in Q1 of 2013. By the end of Q2 Shodan was allegedly [39] capable of scanning the entire Internet once every month. Having scanned all possible IP addresses in Finland, the 61% increase in results is not surprising, although only 8% increase in industrial automation devices is very low.

A closer look at the results given by Shodan reveals one possible reason for the low increase of industrial automation devices: some devices found in Q1 no longer exist in the Shodan database. Shodan claims that no results will ever be erased from its database but clearly this is not the case. It is possible that due to errors or corrupted data some results have been erased. Combining this with the fact

that devices might have been actually taken offline or protected against scans, it is possible the search in Q1 revealed devices that the search in Q4 did not.

Looking at individual search queries in the category for industrial automation in Q1 and in Q4, in some cases the amount of results has declined and in some cases grown, explaining partly the low gain percentage of 8% in industrial automation devices. Figure 7 illustrates how some keywords provided more results and some less, while nine out of the 27 keywords had no change between Q1 and Q4. As can be seen from the figure, only seven keywords resulted in fewer results but two of them did so with relatively big negative numbers, -28 and -49. Similarly on the other end, two keywords provided the biggest increase in the results. If we assume that a decline in Shodan results is not possible, and set all negative sums as zero, we would get a total of 139 new devices, or a 22,5% increase in results.

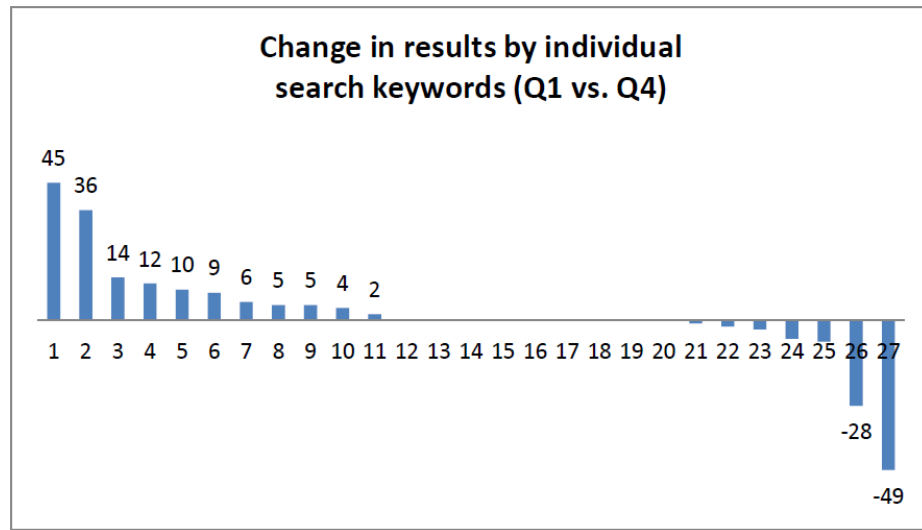


Figure 7: Comparing the change in results for 27 individual keywords used to find industrial automation devices (Q1 vs. Q4, 2013)

3.4 Network traffic monitoring and warning system

In Finland the national cyber-emergency response team (Cert-fi) has deployed, in cooperation with companies in the critical infrastructure sectors, a system [41] to monitor network traffic and warn about ingoing and outgoing malicious activities in the networks of the collaborating companies. The system is called Havaro, a Finnish acronym for a system with monitoring and early warning capabilities. In the end of 2012, Havaro consisted of 12 sensors deployed in the networks of different companies having a role in the critical infrastructure. Havaro-sensors are intrusion detection system devices with optional firewall functionality, and they operate in the border between public Internet and the company's private network. The sensor passively monitors the network traffic in both outwards and inwards directions. Based on trusted sources of IP addresses involved in malicious activities, the sensor alarms if traffic towards those malicious addresses is detected. The sensor also records all

traffic to those addresses and informs a centralized Abuse Helper-process in the Havaros central computer. The Abuse Helper is for event processing: it gathers statistics about the triggered alerts and forwards the information to an analyzer process, which can analyze the captured packets from the triggered traffic flow. A visualization system provides situational awareness data and visualizations about the triggered events. Situational data is being sent back to the customers to keep them informed about the triggered events. Also reports to internet service providers (ISPs) are made, because the ISPs have the best capabilities for blocking malicious addresses.

Havaro is a well working concept but the downside is currently the low amount of sensors. Participation by critical infrastructure companies is voluntary and the customers have to pay the quite substantial equipment cost of the sensor themselves. In return the companies do get enhanced security for their network but they might still feel that the costs outweigh the benefits. Havaro is good for identifying threats passively but in order to provide a complete situational awareness status for the entire nation, a lot more sensors would be needed. Also, from the viewpoint of cooperation of different actors towards better national security, the Havaro system does not deliver much, as Cert-fi is bound by privacy contracts and will not disclose triggered events in the Havaro system to third parties.

When implementing the Havaro sensor into the network of a company, a contract is being made between the company and Cert-fi to protect the privacy of the company's data. The company can decide what kind of data the sensor monitors, and all ownership of information remains with the company. Cert-fi only has the rights to use the data in reporting, visualization and protecting purposes. Cert-fi also has a strict privacy contract with the companies involved making the location of the sensors, and the name of the companies, classified.

3.5 Summary

A lot of exposed ICS devices are in the wild and tools like Shodan make vulnerable devices rather effortless to find. It is clear that especially in the industrial sector securing networks is not at the top of the list of priorities for companies and organizations. Productivity and physical safety of personnel and equipment is ranked well above information security. Nevertheless, as a property of Internet-connected devices, systems can be mapped and scanned by entities other than the owner of the system. Scanning devices for vulnerabilities can be done by external parties, although caution is advised whenever scanning systems in the public Internet for not causing disturbances to the target systems. If the scanning is done in national security purposes, and to disclose vulnerabilities only to authorities or the device owners, the scanning is likely to have a firm ground to stand on when considering legal aspects of the activity, as is discussed further in Section 4.4. A third party entity could scan for devices being part of critical infrastructure to find, report and remediate vulnerabilities, thus improving the security of the national cyber-space. Scanning a nation can provide valuable information when trying to form a situational awareness state of the nation. In the next chapter, an automated system to

identify vulnerable ICS devices inside a nation is proposed.

4 Automated system to find exposed industrial control system devices

In this chapter a system is proposed to detect vulnerable devices in the use of industrial automation networks, such as the critical national infrastructure elements in power and water distribution networks. The system, called KATSE (Finnish acronym for Kansallinen Automaatiojärjestelmien Tarkastus- ja Suodatusjärjestelmä), would constantly scan the entire nation to find exposed, i.e. Internet-connected, ICS devices and analyze them for possible vulnerabilities such as unpatched software versions. Scanning the whole nation provides valuable data about the state of the nation's critical infrastructure, and preventive measures can be taken to secure vulnerable devices. Automated systems providing valuable empirical data, like the proposed KATSE-system, can be very useful in forming an overall situational awareness analysis about the nation's vulnerabilities in the cyber-space.

4.1 Overview of the concept

KATSE is a system that frequently scans the IP address space of the entire nation and tries to identify vulnerable ICS devices accessible through the public Internet. Only devices which are, for some reason, connected to the Internet without proper protection will be found. Some systems are meant to be publicly reachable but some are falsely configured and unknowingly exposed to remote threats. Frequent scanning not only finds vulnerable targets, it also provides the possibility to do in-time analysis of exposed ICS devices to monitor if any action is made to reduce the amount of exposed devices. In the ideal situation where all of the nation's ICS devices are properly secured and not exposed, the most effective scenario for KATSE would be its ability to quickly find new devices appearing in the public Internet and enable counter-measures to be taken before any adversaries can exploit the problem.

While KATSE can be highly effective in finding exposed devices on the Internet, it is not able to detect any devices which are located inside private networks and shielded properly for example with firewalls. Figure 8 illustrates what kind of systems can be found with KATSE. The green arrows show a successful scanning of an exposed ICS device, while red arrows are failed scanning attempts where a firewall is in place to protect the ICS device from Internet exposure.

As the past has shown [32, 42] most devastating attacks come from within a private network from an infected device. KATSE tries to minimize the potential for the public attack vector while other countermeasures need to be taken to secure the private networks from within. For example Finland's legislation in its current state does not allow intrusion to systems even if the purpose would be to do security auditing to improve national security. To assist in making necessary changes to improve system security, KATSE could in theory also be used to penetrate networks with flawed security and do vulnerability analysis on the devices inside a private network. However this raises a lot of legal and privacy concerns, which are both explored later in Section 4.4. "Legal issues".

A nation-wide scanner gaining vulnerability information can arouse suspicion

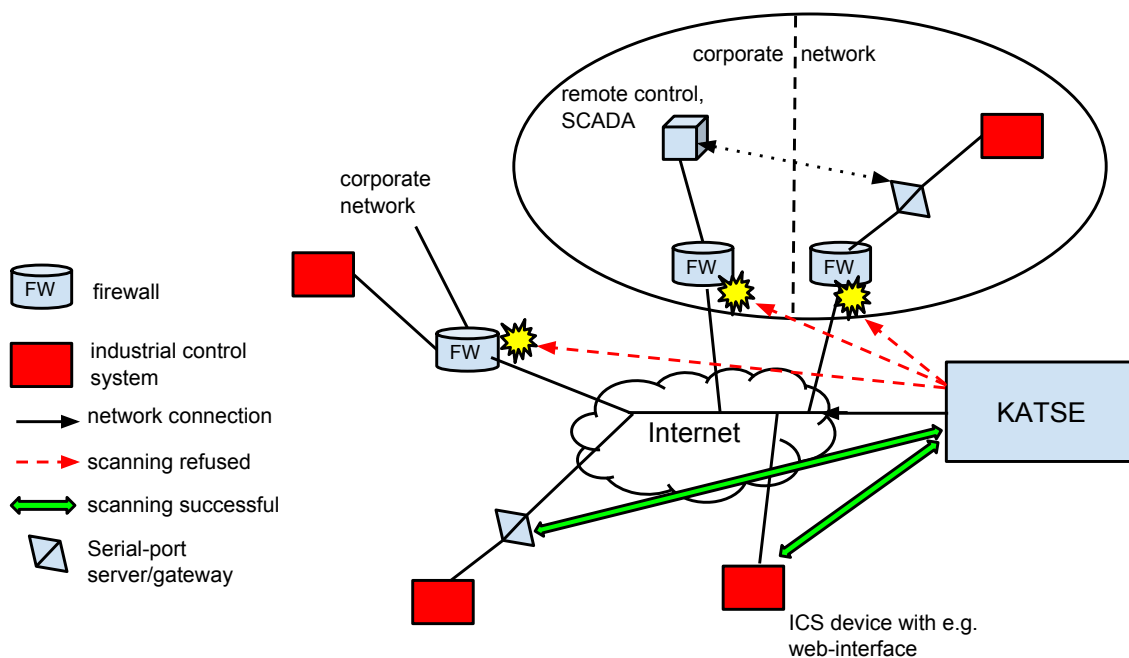


Figure 8: Illustration of the systems KATSE can find.

among companies and the public. To induce trust, KATSE could be run by a trusted entity: authoritative organization, or an organization approved and appointed by authorities. In Finland Cert-fi, Criminal police (KRP), Finnish Security Intelligence Service (SuPo), Finnish Defense forces and the National Defense University (NDU), and National emergency supply agency (Huoltovarmuuskeskus), would be likely candidates. Cert-fi is already running the Havaro-program to help critical infrastructure companies to defend against malicious traffic, and KATSE would probably fit nicely into their operations. CERT also has the incentive and know-how to produce and operate KATSE. From police entities, KRP and SuPo, security of the society is one of SuPo's objectives along with counter-intelligence anti-terrorism activities. However, SuPo is more of an operational entity and KATSE would not fit into their main objectives. The same goes for the Criminal police; KRP focuses on criminal activity, and would not probably be interested in such system that only points out vulnerabilities and does not monitor or provide evidence against attackers.

In the time of peace the Finnish Defense Forces do not have jurisdiction, but the NDU is interested in a comprehensive cyber-space situational awareness system, and KATSE could provide valuable information to their system. NDU would not probably be interested in running KATSE, but no matter who is running the system, information from KATSE should be provided to NDU to help the forming of situational awareness about the nation's cyber-security. National emergency supply agency's objective is to prepare for emergency situations and to ensure the critical functions of the society stay functional even in the time of crisis. KATSE would definitely be in assistance for helping the nation to proactively protect important assets. However, the agency probably would not want to run the system; they could

rather oversee its usage and results while leaving the operational functions to for example Cert-fi and aid the operations financially.

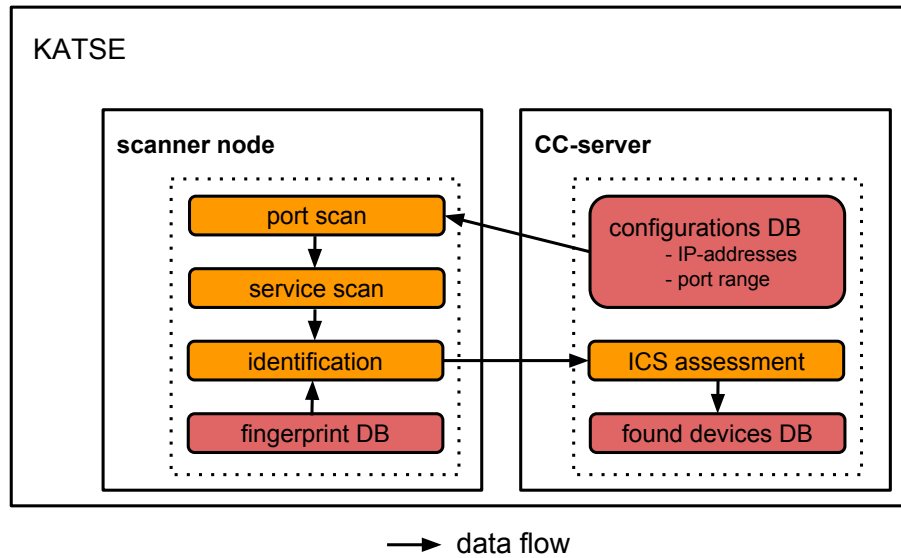


Figure 9: Important components of KATSE

KATSE consists of scanner nodes and a central command and control server (CC-server), as can be seen in Figure 9. Scanning is done non-intrusively and sourcing from the public Internet, finding only devices connected directly to the Internet. That way no private systems are breached. The CC-server passes on configurations for each node, including IP addresses and port range to scan. The scanner queries the given ports for all the given IP addresses to learn, which hosts are online and what ports are found to be open. After the port scan, a more detailed service scan is initiated towards the hosts in order to learn identifiable data from the target device. That data is then compared to an ICS device fingerprint database in the identification process. If a positive identification is made, all the data about the suspected ICS device is sent to the CC-server for further analysis in the ICS device assessment process. Information about all the found ICS devices are kept in the CC-server, and can be delivered to authorities, like national CERT, to enable further actions in securing the exposed devices.

Next in this chapter the concept of proposed KATSE system is presented in more detail, explaining the overall architecture of the system, methods for scanning, how targets are analyzed, impact of KATSE on the national network traffic, and evaluations about the possible short-comings of the system are presented. Also one section is dedicated to cover legal concerns when running a system like KATSE. Chapter 5 describes a proof of concept prototype of the system with reduced functions.

4.2 Scanning

Scanning tasks are distributed for multiple nodes, and each node in the system is responsible for scanning a given set of IP addresses over and over again. The scanning is done in two stages: first to quickly go through the assigned address blocks to find online hosts and open ports (port scan), and on second stage (service scan) doing more specific service queries depending on which ports were found to be open. Technical details how to implement the scanning tasks are not discussed, instead the focus is to explain important ports and services and what information needs to be captured to make the concept viable.

The nodes start the scanning operation by doing an initial scan for the IP addresses which are allocated to the node. For each online host, a port scan is done querying the desired ports. This port scan is meant to reveal ports that are open on the target host. The second stage of scanning connects to open ports and collects the response data from the hosts. This separation into two stages is done to save time and network resources. As finding up hosts and open ports can be quite fast, establishing a connection with the target host and applying various scan probes is more time consuming and should only be attempted on open ports. More on time and traffic consumption is explored later on in this chapter.

Gathered information and important port numbers

As there are overall of 65335 ports standardized for computer systems, scanning every possible port on every target host would be really time consuming. Scanning only desired ports improves the performance of the system and provides less useless data. From previous research and from sources covered in chapter three [4, 38, 37], a list of useful ports is presented in Table 2 below. The table starts with ports that are used by generally popular protocols and the ports from 135 forward are used by ICS protocols. Notes are attached to explain them further.

Open ports are queried and header-responses and payload information coming from the host is recorded. It is later explained how the information is used to analyze the target, but in short, the gathered data is compared to an ICS device fingerprint database to determine if the target is of interest or not. In other words, scanned information is compared to existing fingerprints: if a positive match is found, all the learned data is sent to the command and control server for further analysis. Below, it is explained what kind of information certain protocols can reveal.

- FTP and Telnet: headers can reveal system and software information
- HTTP: open HTTP and HTTPS services are queried for their HTTP-header and HTML-title. Header can reveal software information, like used web-server software, and HTML-title extracts the title of the hosted web page, which can reveal the purpose of the web server such as a remote control interface over the web.
- SNMP: can reveal information about the hardware, such as device name and model, and possibly host names from the local area network.

Table 2: Useful ports for analyzing target devices

Port	Protocol	Notes
21	FTP	-
23	Telnet	-
80, 8080	HTTP	-
137	netBIOS	-
161	SNMP	-
443	HTTPS	-
3389	RDP	Remote Desktop Protocol for Windows
135	OPC	protocol used in ICS environments
502	Modbus/TCP	serial bus protocol over TCP
771	RealPort	Proprietary protocol used in serial-port servers and proxies
2001-2010	Various	Default ports to some serial-port servers
3001-3010	Various	Default ports to some serial-port servers
4840	OPC UA/TCP	Protocol used in control systems

- netBIOS: can reveal system information from the private network of the target host. Can reveal MAC addresses, host names and user names.
- OLE for Process Control User Agent (OPC UA): usually needs specific software and authentication to communicate with OPC UA service. Open port suggests that a control system device might reside in the target address.
- Modbus: can confirm that Modbus/TCP used in ICSs is running on the target device
- 2001-2010 and 3001-3010: responses can confirm that serial-port server is running on the target device. Device can usually be interacted with since security is often weak. Also port 3002 is sometimes used for a Telnet service in Windows CE operating systems used in ICS devices.
- RealPort: reveals that a serial-port server is running in the target address.

In the research for exposed ICS devices in Finland [4], Telnet, HTTP and SNMP protocols were the most useful in identifying hosts. HTTP-header often gave away used web server software, which was very useful in identifying more similar ICS

devices, as many ICS device vendors use their own web server software differentiating them from common use web servers. On some occasions Telnet-protocol gave away the used system by revealing too much information in the command prompt, even before authentication was requested. In many cases SNMP-protocol responded with the devices name, and sometimes with software and firmware versions as well. Getting the specific name and model of the device with firmware version, it is easy to find out in what kind of environments the device can be used, and whether the device has any known vulnerabilities to exploit.

In Figure 10, a couple of examples taken from Shodan [15] show what kind of information the protocols might reveal. In the first example, query to port 23 using Telnet-protocol reveals the name of the target device, the MAC-address and even the firmware version. That particular MAC-address resolves to a company called Moxa Technologies, which manufactures third-party control and gateway devices for industrial and building automation sectors. The MAC-address in the second example, given out by the NetBIOS-protocol, reveals even more information: resolving the address reveals that the device behind the scanned IP address is manufactured by Siemens AG A&D ET, where A&D stands for automation and drives, and ET for electrical installation technology. This way MAC-addresses can be used in identifying devices, or in determining their purposes.

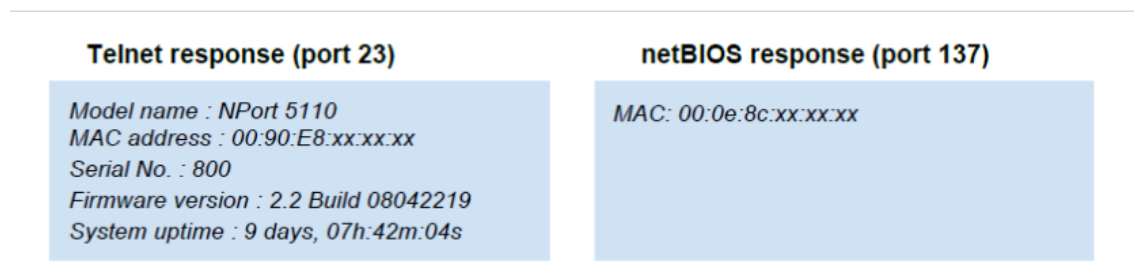


Figure 10: Responses from scanned devices reveal sensitive information

Additionally, getting HTTP-payload data from ICS devices with open web servers can provide valuable information. This payload data is the actual content which is hosted at the web server, and in this case, the content of the front page of the site. However, as later explained, to reduce traffic the HTTP-payload is not automatically captured on every device having open HTTP-ports, so the payload information cannot be used in the device identification process. Only after identification, target hosts recognized as ICS devices, will be queried for the payload data if an open web server is detected in the host. The payload will be stored in the records along with other data from the specific device, and can be later examined. Examination of the payload data automatically is possible but would need a fine tuned system to detect any useful data. Existing payload data in the records can, however, help if a human analysis of the target is being conducted at a later stage.

Taking a screenshot of the web page that the target device is hosting is one method for additional information about the target. Although a screenshot gives essentially the same information than the HTML-page decoded from the HTTP-

payload, it would be a quicker way to get more information about the purpose of the device when human interaction with the records is assumed. Also for visualization purposes a screenshot can be more convenient than an HTML-page.

Some information about the target host can be gained without communicating with it. The IP address of the target can reveal the owner organization of that address, and in some cases, give approximate geographical location information. All this information depends on who-is databases maintained by the Regional Internet Registries (RIR), which are responsible for assigning IP addresses to Internet service providers. ISPs are responsible for updating the who-is database by adding name, organization, location and contact information of their customers. However, ISPs are not required to add details about every customer, and they generally add only organizations to the database. The importance of the IP address, owner information and location information is discussed in more detail later in the chapter.

4.3 The concept in detail

To get the idea how a concept like KATSE could be built and what needs to be considered when planning such a system, details of the overall structure and operations of the system are presented in this section. First, the main elements of the system, scanning nodes and a central server, are introduced. The process of analyzing the found ICS devices is explained, and later on practical matters such as induced traffic, time constraints, physical location of the nodes and possible problems detected in the concept, are covered.

4.3.1 Overall architecture of the system

As the figure 11 illustrates, the system consists of several nodes doing the port scanning and a central command and control server. The nodes can be geographically distributed and, besides scanning, the nodes will communicate only with the CC-server using a virtual private network (VPN) to secure the communications. Also a VPN should be used when connecting remotely to the CC-server for management tasks.

4.3.2 Components

The nodes are computers with software capable of doing the required scanning tasks. Each individual node has the scanning software and an ICS device fingerprint database. Scanning software is used to get fingerprint data from services running in a given IP address and the data is then compared to the ICS database. Positive match concludes the work of the node concerning the specific IP address: the node sends all acquired data from the positive match to the CC-server for further analysis. Having a fingerprint database on each of the nodes greatly diminishes the need for communication between the nodes and the CC-server, saving time and bandwidth. By giving the detailed analysis task to the CC-server, the scanner will remain effective in scanning the assigned IP addresses. Since most of the scanned IP addresses do not provide positive matches, it would be unfeasible to send all scanned data

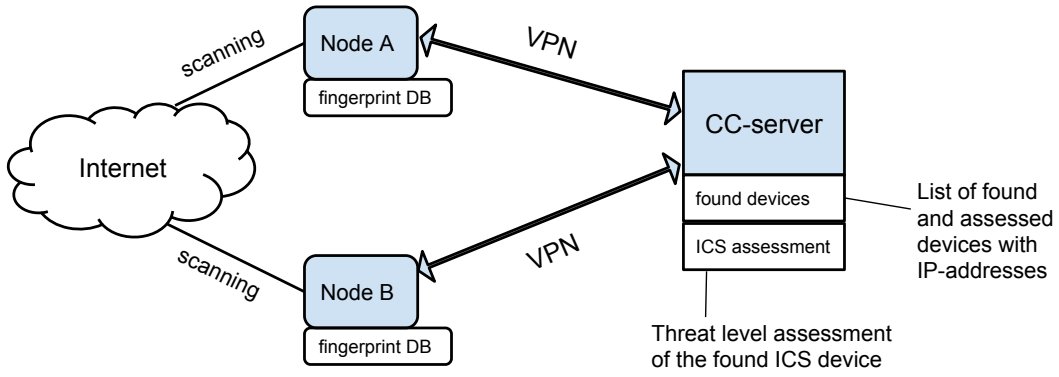


Figure 11: Components of proposed KATSE system

to somewhere else for identification. The analysis-part of positive matches could also be done in the nodes but having a central server makes collecting data, keeping databases and managing the system easier.

Database of industrial control system devices

The private database of ICS device fingerprints is present at every node and additionally in the CC-server for managing the database. The database includes known fingerprints for industrial grade devices used mostly in different kind of industrial processes, control systems and industrial automation systems. The fingerprints can be gathered from multiple sources. Public sources like Shodan and Metasploit Framework are useful along with private scanning efforts. The fingerprint entries in the ICS device database should include data from different services typical for a specific device, such as an HTTP-fingerprint, a Telnet-fingerprint and an SNMP-fingerprint. The database should include as many fingerprints from a single device as possible to increase the chances of positive identification. TCP/IP-stack fingerprints, gathered e.g. by Nmap, could also be added but currently the Nmap service discovery, which makes use of the TCP/IP-stack of the target software, is quite slow for purposes of scanning huge amount of hosts. Nevertheless, TCP/IP-stack fingerprints can be included in the database for additional information even if TCP/IP-stack discovery would not be included in the primary scanning scheme. The database should be updated whenever new device fingerprints are acquired or when new devices need to be found, and the CC-server is responsible for passing database updates to the nodes. An example of the ICS device database is presented below in Table 3.

Along with ICS device names and fingerprints, the purpose of the device is explained shortly to get an initial idea what the device is capable of doing. When ICS devices are found, a report is constructed and the purpose statement is attached to the report to make analyzing the target device easier and faster. Information about the purpose should come from a reliable source such as the manufacturer of

Table 3: Example of information contained in the ICS device database

Device name	HTTP	purpose
Siemens Simatic S7 CP (CP 343-1 CX10)	HTTP/1.0 302 Location: /Portal0000.htm	designed for use with SIMATIC S7-300 automation systems. It enables the connection of S7-300 PLCs to Industrial Ethernet and PROFINET IO via 2 ports
Schneider TSX Micro ETZ510-1	Server: Schneider-WEB	communicates with Micro-PLCs, part of Schneider automation platform

the device.

Command and control server

The command and control server is responsible for doing the vulnerability analysis of the found devices. When a node makes a positive identification of an ICS device, the node sends the scan data to the CC-server, along with the name of the device which was identified. Vulnerability database is then queried to find if any exploits exists for the ICS device. The server tries to get more information about the purpose of the device by querying the device's IP address of its owner and geolocation data. The overall ICS devices threat-assessment process explained on the next section, should be done on the CC-server to save processing power on the scanning nodes.

Operations of the server are automated but remote management can be used to update databases, update IP address allocations, and extract data and so on. Also additional scanning tasks can be forced remotely to get more detailed information, like the HTML-payload data, from a specific target.

4.3.3 ICS Device assessment

The scanner is expected to provide a lot of results, especially at first after starting the scanner. The passive scanning methods of KATSE will identify target devices of different importance. Two ICS devices identified with the same fingerprint might be used in totally different environments; the other could be a PLC controlling a manufacturing facility producing car parts, and the other might control water flows in a local waste water treatment plant. Distinguishing important, critical devices from non-critical devices would be important to quickly remedy the most critical vulnerabilities in national infrastructures.

The point of ICS device assessment would be to automatically estimate the importance of the device and its vulnerabilities, and present an overall threat-level assessment for each device. Without any automation of the process, a human involvement is needed to go through every found device and analyze the target for its importance, and also for detecting false positive identifications. One option to

exclude human estimations from the process would be to report every device found straight to their owners based on the IP address of the device. However, that approach probably faces problems with the recipients ignoring the matter and possible getting frustrated of constant reports of exposed devices. After all, the recipients are responsible for the devices being exposed in the first place.

The automatic assessment of the devices is realized to be very difficult and sensitive to errors. The process would have to be fine-tuned to avoid wrong interpretations and it should be trustworthy to a certain extent. Such a system would need a lot of research; some ideas are presented here of what information could be used in the assessment process. Both existing data about the device prior to scanning and newly gained information from scanning such as IP address, open services and raw scan data should be used when assessing the device. For example, the assessment process could take advantage of items presented below.

1. Purpose: based on the brand and model of the found device.
2. Importance: keyword analysis (owner, domain name, HTML-title, FTP-title, payload)
3. Vulnerability level: private and public vulnerability databases can be referenced for software vulnerabilities and existing exploits.

In Figure 12 a chart is presented to show how the automated device assessment process could work. The device assessment is triggered after a scanning node makes a positive identification of an ICS device and sends related data to the CC-server. The CC-server then parses and uses the received data; three key assessments (marked in blue) affect to the overall threat level (marked in red). The IP address and payload data from scanning such as HTML-title are used as data in the keyword analysis process producing an importance value. The ID of the device is used to get a purpose value from the ICS device database. Vulnerability analysis produces a "1" or a "0" whether an exploit exists for the device or not. A final report (marked in green) is constructed based on the assessment process and the report is stored locally. If the threat level is high enough (e.g. medium/high) the report is sent forward to an administrator along with an alert that a potentially crucial ICS device has been found.

Purpose of the devices, just like fingerprints and known vulnerabilities, is information known prior to scanning. The purpose-factor is mainly related to the capabilities of the device. Expensive, robust devices designed for heavy-duty industrial usage are considered more important than for example cheap gateway proxies. The true importance naturally depends on where the device is actually implemented, and that information is usually hard to get. Gateway devices can reside in very important critical infrastructure systems or just at a local supermarket enabling remote supervision of cooling thermostats.

The keyword analysis producing an importance value takes use of information gained during the scanning. The IP address of the device might be registered to a

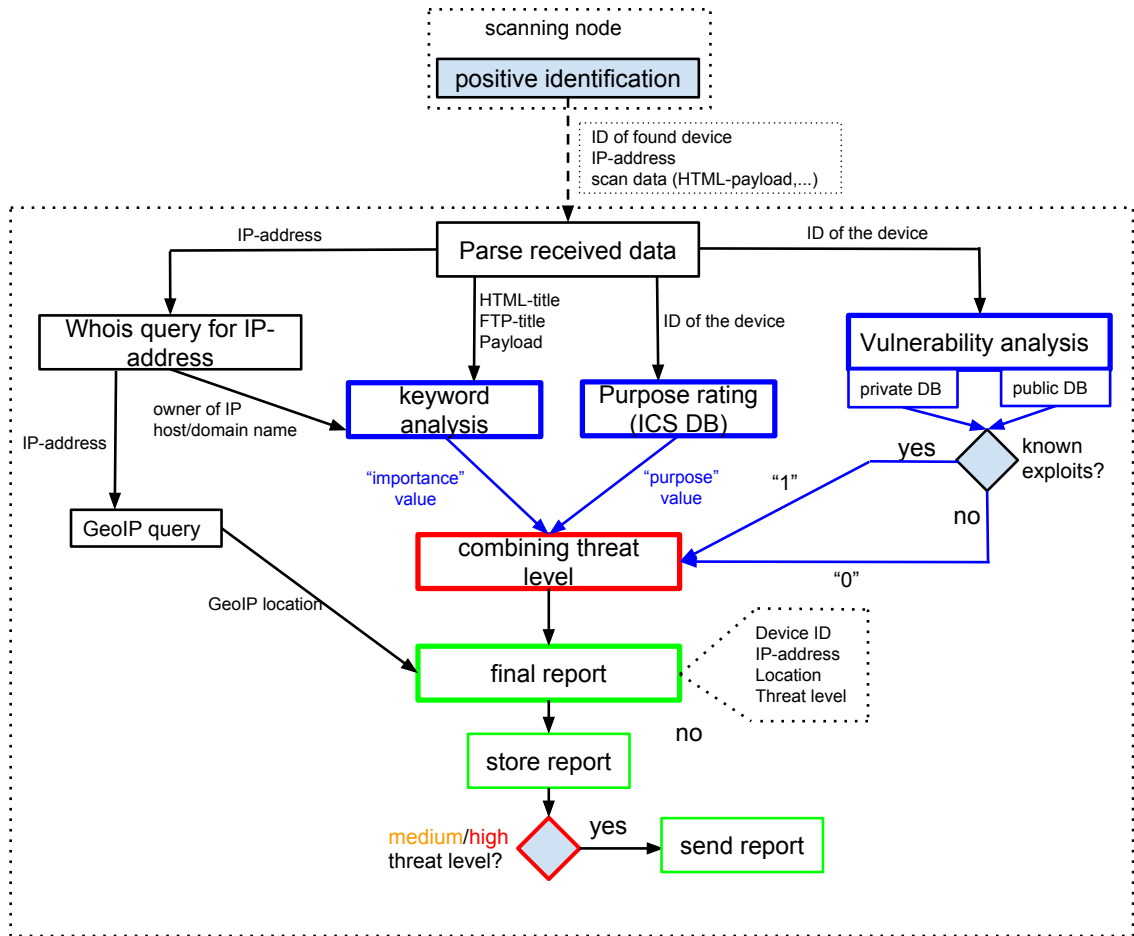


Figure 12: ICS device assessment process for devices found by scanning

known industrial company, or the HTML-title of the web-page hosted in the target device might reveal important information such as the physical location of the device. HTML-payload of the web-page can further reveal the purpose and capabilities of the devices. Also passiveDNS databases can be referenced to find out if the IP address has previously had revealing domain names indicating an industrial system. PassiveDNS databases are hosted by domain name server operators, and the databases collect passively every domain name which is associated to a certain IP address. As stated, the overall analysis process would need a lot of work for making it effective and reliable, and that would definitely be one goal for the future for improving the KATSE-concept.

4.3.4 Practical issues of the concept

This section describes practical matters concerning the amount of needed nodes and induced traffic from the scanning. The practical issues are considered if a system like KATSE would be used in Finland. The amount of IP addresses in Finland is discussed and later, scanning speed and traffic amounts are covered. The latter

answers an important question whether a system like KATSE produces significant traffic volume compared to the overall network traffic in the Finnish communications infrastructure.

According to the database of MaxMind [43], Finland has 13 672 280 unique IP addresses allocated. This is approximately the maximum amount of IP addresses that can be used inside Finland, if all currently assigned IP address blocks would be fully used. The actual number of addresses in use, however, is a lot lower than the maximum amount. In order to serve new customers, ISPs have reserves from where to assign IP addresses to new customers. Similarly organizations might purchase excessively large address blocks from ISPs to support growth and future needs. In addition, the amount of IP addresses does not tell the whole truth about the amount of Internet connected devices, as many devices are located in private networks and share the same external IP address.

The speed of scanning

The speed of a scanning software depends mainly on the software implementation itself and the speed of the used network connection, as a paper [35] published in August 2013 demonstrates, describing a new approach for quickly scanning large amount of IP addresses. Using a new scanner software called ZMap, they showed how the bottle-neck for increasing scanning speed is currently the available network connection, if a regular network interface card (NIC), processor and RAM are assumed. They demonstrate how a moderately priced computer can utilize fully a high-speed 1 gigabits per second network connection when sending out scanning probes with the ZMap. At the maximum rate, ZMap was able to send out 1.4 million probes per second, achieving a 1300 times faster speed than Nmap, a highly popular network scanner. SANS Institute also compared Nmap against another fast scanning software called Scanrand in their white paper [44], achieving approximately six times faster scanning speed with Scanrand than with Nmap. These results clearly suggest that the scanning method of asynchronous scanning implemented in ZMap and Scanrand is far superior to the traditional method of synchronous scanning, which Nmap uses. According to both of the papers, the approximate accuracy of the scanners was similar, even though the scanning speeds were in different proportions.

The biggest factors for ZMap's superior speed against Nmap are well implemented asynchronous scanning and Zmap's ability to bypass Linux kernel when sending out network probes, i.e. packets. Asynchronous scanning means that sending the probes and receiving responses from the hosts are done in two separate processes. One process only sends out probes as fast as it can and the second process catches the responses. Nmap is able to send out multiple probes at the same time but only to a certain limit: the same process must wait for the probed hosts to respond or to receive time-out messages before sending out more probes.

While a regular ZMap implementation is promised to scan the entire Internet in less than an hour, it is good to note that only one scanned port is assumed, and increasing the amount of scanned ports to 10 would probably make ZMap 10 times

slower due to the nature of the software. Both the ZMap and the SANS paper ran the scans to only one port on each host, and as explained, scanning for more than one port is essential for KATSE. Tests ran in a private network revealed that Nmap performs better when port number is increased: scanning a 256-address local network for 10 ports took 5,7 seconds and scanning for 20 ports took 5,9 seconds. The effect is quite small, so the increased scanning time should not be a key concern since more scanned ports can reveal more information from the target IP address. If considerably more ports would be scanned time would become an issue. Scanning a single host in a local area network for 10 ports took 2 seconds, 1000 ports took 5,1 seconds, and scanning all possible 65535 ports took 111,2s seconds. Scanning all possible ports for a large amount of IP addresses is clearly unfeasible, and scanning only tens of carefully selected ports is fast and can give a fairly good overview of open ports in the target host.

To sum up the speed of Nmap, Scanrand and ZMap, the time to scan all IPv4-addresses in Finland is calculated. The speed of Nmap and Scanrand are taken from the SANS white paper [44] and the speed of ZMap from the ZMap paper [35]. Time to scan 13 676 636 IP addresses (amount of IP addresses in Finland according to MaxMind database [43]):

- Nmap: 2697 minutes (45 hours)
- Scanrand: 445 minutes (7,4 hours)
- ZMap: 10 seconds

Like mentioned earlier, no software proposition about the implementation is made. Comparing the speed of the scanners reflects on what kind of a time frame is realistic to scan an entire nation. It is good to note that while ZMap is very fast in port scanning, the slowest part of the scanning in KATSE is presumably the second stage, where open ports are queried and responses recorded. For this purpose Nmap, for being a long developed open source scanner, has already plenty of functions available. ZMap with some changes could be a lot more efficient in gathering information than Nmap but its potential is still a bit unclear. The issue with ZMap is the huge amount of bandwidth it is able to consume, as later explained.

Comparing different protocols, getting HTTP-headers from web servers is among the biggest efforts traffic wise. To get an estimate of the speed of the second stage of scanning, HTTP-headers of 10 Finnish web servers were fetched with a Python script. The average rate of 7,9 hosts per second was achieved in a test of 10 consecutive scans to the same 10 web servers. The bottle-neck in the test seemed to be the response time of the queried server, as was expected. Some servers were always a lot slower to respond than others, response time varying on average from 1 millisecond to even 400 milliseconds. With a fast Internet connection adding the functionality to grab also the HTTP-payload from the front page of the hosted site did not greatly increase the execution time of the script, although as seen in Table 4 below, traffic amounts were a lot higher. Querying the same 10 web servers and grabbing also the payload from the front page, the script performed on average at the speed of 6,45 hosts per second.

Table 4: Summary of measured scanning times

	average time of 10 runs (s)	traffic
10 web servers without payload	1,27	100kB
10 web servers with payload	1,55	1430kB
single host, port scanning 10 ports	2,0	-
single host, port scanning 1000 ports	5,1	-
single host, port scanning 65535 ports	111,2	-

Traffic

Traffic volume depends highly on the amount of open ports in the target host. The scan tests of five hosts for 10 and 20 ports generated 17,3 kB and 21,0 kB of traffic, respectively, both having the same amount of open ports. Little differences scale up when scanning huge amount of hosts but not enough to justify dropping interesting ports from the scanning scheme.

With Nmap, the first stage of scanning, host and port discovery, would produce approximately a constant traffic flow of 24,3 kbps. That equals to a total of 0,26 GB of data transferred per day per scanner. If a ZMap instance would be used, by default it utilizes practically all of the available network bandwidth up to one gigabit per second transfer rate. With ZMap, the available bandwidth should be carefully allocated to have a suitable amount of traffic. A bandwidth of approximately 855 kbps would be enough for ZMap to be at the same scanning speed as Nmap. Definitely this would not be the ideal use case for ZMap which is built to maximize scanning speed.

Querying web servers, which is part of the second scanning stage, produced a constant traffic flow of 638,3 kbps, when grabbing just the HTTP-headers from the server using a python script. That equals to 6,9 GB of traffic per day per scanner. If also the payload from the web servers would be grabbed, it would realize over tenfold amount of traffic being transferred, approximately 7400 kbps to make a total of 80 GB traffic per day per scanner. From the perspective of traffic and time, it does not seem reasonable to do payload-grabbing as a default to every web server. Instead, it could be used as an additional function for only positively identified ICS devices.

HTTP-headers are among the largest responses the scanners will receive, so the size of the headers helps to estimate the total traffic amount KATSE would produce. When adding the discovery stage together with the HTTP-traffic, the scanner would produce a total traffic flow of 662,4 kbps. Compared to the overall amount of traffic in the Finnish Internet Exchange points in Otaniemi, Pasila, and Oulu (FICIX points), which pass through the majority of Finnish Internet traffic, the traffic produced by the scanner would be almost unnoticeable. A yearly average for the total amount of traffic going through all three exchange points is 18,6 Gbps (stats.ficix.fi). Even if 10 scanners would be in a constant use, they would still account for only 0,0004% of the Finnish Internet traffic. Adding the fact that not all traffic in Finland goes through the exchange points, it is safe to say that the traffic flows produced by the scanners would not be a burden to the overall infrastructure.

Comparing a ZMap instance, scanning at the maximum speed of 1 Gbps, to KATSE, ZMap would constitute quite a significant load on the infrastructure, little over 5% of the average traffic volume in Finland. Traffic volumes are summarized in Table 5.

Table 5: Summary of traffic volumes compared to FICIX

	Traffic/s (kbps)	Traffic/day (GB)	% of FICIX
Nmap port scanning (10 ports)	24,3	0,26	0,0000013%
Grabbing HTTP-headers	638,3	6,9	0,000034%
Grabbing HTTP-payload (front page)	7400	80	0,0004%
FICIX (yearly avg.)	18,6 Gbps	200,9 TB	100%

KATSE as a concept does not require a high speed scanning scheme. Scanning through the entire nation once a day is fast enough, considering the passive nature of the system. A few nodes with Nmap software would be sufficient to daily scan through the Finnish IP address space. Considering the traffic loads, the physical location of the nodes is insignificant but placing the nodes to different servers in different networks would increase redundancy of the system in case of hardware or network failures.

4.3.5 Geolocation

Regional Internet Registries (RIRs) assign IP addresses to Internet service providers (ISPs) under their, usually continental, jurisdiction. The RIRs keep a who-is database containing information about the assigned IP address blocks. The ISPs, who assign IP addresses to their customers, can add names, e-mail and street addresses, and contact persons about their customers to the who-is database. Services providing geographical locations for IP addresses (GeoIP or Geolocation) are usually based on the RIR databases, and as a secondary source, the geolocation services might use data obtained with data mining, collecting data from users and buying customer data from third parties. As IP addresses get resigned and they do not contain any location information on their own, the geolocation data and services must be treated with certain ambiguity. [45]

MaxMind, a well-known geolocation service provider, claims to detect the correct country of an IP address with 99,8% accuracy. Within the USA, the correct state is recognized with 90% accuracy, and the correct city with 83% accuracy. For other countries, the accuracy varies: for Finland, MaxMind claims 60% accuracy to provide the correct city of the IP address. [46]

Even though a high accuracy is not possible with geolocation, it is still better for visualizations than not having any information about the physical location of an IP address. With geolocation, found and vulnerable ICS devices can be mapped to cities to give an approximate oversight where the problematic devices are located.

4.3.6 Additional functions

Additional methods for KATSE, which could add value to the concept, are proposed here. These additional methods would not be conducted automatically, but only on-demand basis. Some of the functions are not legal in the current legislation of Finland, however, they are still explored as changes in legislation can be made.

On top of the automatic functions of the system, KATSE could provide scan initiations remotely, in the case of certain IP addresses needs to be immediately queried. Remote scanning requests could also include taking a screenshot from hosted web-sites, or to do a more detailed scanning on the target system. The detailed scanning could include intrusive methods such as trying default credentials or exploiting a severe security vulnerability on the device. Such actions would undoubtedly prove useful in assessing the devices vulnerabilities or figuring out its purpose, but such actions are currently illegal. The next section discusses legal issues, concerning scanning, in more detail.

As a by-product of scanning an entire nation on a daily basis, KATSE could also collect statistics. While getting the data about found ICS devices is important, trend data of the nation's exposed systems can be useful in demonstration purposes, spreading awareness and as empirical data for decision makers. KATSE could also record data which is unrelated to ICS devices. As the nodes go through all of the Finnish IP addresses once a day, up to date information about the amount of currently online hosts, which ports are most frequently open and so forth, could be recorded. Trend data about the changes in those categories mentioned can be interesting when observing whether, for example raising security awareness, has had any real impact.

4.3.7 Possible problems in the concept

This section discusses possible concerns and problem areas of KATSE. As a scanner system, the usefulness depends greatly on the system's ability to identify vulnerable ICS devices. Identification process is based equally on two factors: the ability to get useful information from the target hosts, and the coverage of the ICS device fingerprint database. Devices without pre-recorded fingerprints cannot be identified, and devices with only one, e.g. vague HTTP-fingerprint, will be hard to identify. Only when scan information collides with the database information, a positive match is found. This leaves the system only as effective as the amount of fingerprints in the database. A solution for improving the database is to constantly gather more fingerprints for more devices, in conjunction with vendors, companies in the ICS sector, and security researchers.

A practical issue with KATSE is actually securing the found devices. The owner of an IP address cannot always be resolved, and in any case might require cooperation from the ISP of the target device. ISPs might not want to disturb their customers and disclosing customer information to a third party is probably forbidden. Even in the case of the owner of an exposed system is found, no guarantees can be made: the owner might not acknowledge the problem or is not simply willing act on the matter. For the industrial sector to realize the importance of information

security, awareness needs to be raised about the existing threats along with ways to protect the systems.

Disturbing scanned devices or networks behind them

When scanning IP addresses, a certain level of interaction with the target device is done. Communications consume processing power and bandwidth from the target device and network. Nevertheless, because the scanned device is connected to the public Internet, it must be able to handle queries by basic communications protocols without problems. Just for being Internet-connected, the device probably receives a lot of queries around the globe from scanners, bots, spiders and other random sources. While a large amount of data intensive connections can be really harmful for an embedded device with low CPU power, simple scans, like the type which KATSE does, will not likely cause any issues for the devices. Industrial automation systems might have even ping-sensitive devices [6] inside the control network but those devices would not be directly connected to the Internet, rather than reside behind gateway devices. Only by using those gateway devices as entry points to target the sensitive devices behind them, would serious harm be possible. While KATSE would not be doing this, harmful actors might be, and that is one reason to locate such gateway devices and make them safe before malicious actors can cause harm through the gateways.

Analyzing target attractiveness

One key concern, along with the system's ability identify devices, is to its' ability to analyze the importance of that device. False positives, when a non-ICS device is identified as an ICS device, can be kept low with strict rules for identification through correct fingerprints. Gathered information can be so scarce that even when an ICS devices is found, determining its' importance can be impossible. For this reason the system can be given certain parameters which classify the found devices into importance categories but leave further estimations to authorities receiving the vulnerability report. In the worst case authorities might receive a lot of vague vulnerability reports containing only the name of the, purpose extracted from the ICS device database, and IP address with only country-level geolocation information available. It would then be up to the authorities to manually check the target out by contacting the owner of the IP address and assess the importance of the device. This can be by all means very ineffective use of human resources. That is why only reports with high enough threat level should be automatically forwarded to authorities.

Inadequate location information also makes the analysis harder, especially when trying to visualize the overall status of the nations vulnerable ICS devices. Internet service providers would have more precise knowledge of the whereabouts of a certain IP address, but they are bound by privacy laws not to give customer information to third parties.

False negatives

The amount of false negatives in a system like KATSE can be large. False negative means that the system was not able to identify the device as an ICS device, even though it was one. If the target device is Internet-connected but does not provide any unique or identifiable fingerprints rather than just general fingerprints, the device will not be detected as an ICS device. This means that also harmful actors trying to locate such devices will have a hard time identifying them. On the other hand, other clues such as the owner of the target IP address (e.g. an energy company) might be enough for an adversary to try intrusive techniques to break into the system and gain more information. Intrusive actions, where the device protections are breached, are forbidden by the Finnish legislation thus making KATSE unable to make the same kind of attempts in gaining information than adversaries might do. Improving the ICS device database, adding more fingerprints and improving the scanning will help to reduce the amount of undetected ICS devices.

4.4 Legal issues

Deploying and using a system like KATSE comes with certain legality concerns. Legislation in different countries varies a lot and for each country the legislation should be reviewed individually when a system like KATSE is considered. One problem in cyber-security is setting the boundaries. While criminals exploit systems at will, researchers and authorities must obey the laws when conducting research, analysis and investigations of cyber-offenses. The police can have additional rights through search warrants to investigate a suspect but that does not help in preventing crimes. In the USA, national agencies, like NSA and FBI, seem to have permissions to widely monitor communications networks but that is not the case for example in Finland. Authorities in Finland do not have the permission or the resources to analyze web traffic to the extent that NSA does, to prevent serious threats.

Systems like KATSE can be valuable assets in preventing attacks or at least lowering the amount of attack surfaces in a nation. It is in the nation's best interest to have a legislation that allows security research and analysis with good intentions. This section briefly looks into the Finnish legislation about computer and communications security and privacy, estimates the legality of KATSE in Finland, and explores the possibility to do intrusive scanning from a legal viewpoint.

4.4.1 Finnish legislation related to communications

In the Finnish criminal law, chapter 38 covers communications offenses. From our point of view relevant articles are from §5 to §8. §5 and §6 criminalize disturbing or preventing communications. §7 forbids disturbing or preventing any computer system from working with manipulation of data (sending, transferring, harming, altering or removing). Also an attempt to do any the above mentioned is punishable. §8 covers breaking into computer systems. Breaking into a system using credentials not belonging to the user is forbidden, as is breaking in by otherwise circulating

security measures. Also getting information from a system with a special technical device is forbidden, as are attempts to do either of the acts above.

Data manipulations and denial of service attacks fall into the disturbances mentioned in §5 and §7. As §8 states, trying default credentials or guessing passwords without the permission of the owner of the system, is forbidden by the law. The same article also criminalizes the use of any exploits to gain access to a system, no matter how weak and vulnerable the system might be. The second moment of §8 forbids gaining information from inside a computer system without actually breaking into the system. It is unclear what the "special technical device" stated means.

Criminal law chapter 34 §9 forbids endangering computer systems. Also, in the intention to cause harm or damage, it is forbidden to import, manufacture, sell or distribute devices or software or code that is designed to endanger computer systems or communications systems. The same applies to breaking an encryption or other protection methods of a computer system. Possessing access codes or credentials belonging to someone else is also illegal. This section of the law raises questions about exploit tools such as Metasploit Framework, which has a huge library of exploits and automated functions to exploit vulnerable systems. Of course, as the law states, using software like Metasploit is illegal only if the intention is to do harm or damage, including hacking and exploiting systems. Metasploit is marketed as a tool for penetration testers and other security enthusiasts, and the developers of Metasploit do emphasize that penetrating systems without permission is illegal.

Intentions matter

During the research on exposed ICS devices in Finland [4], Sari Kajantie from the Criminal police of Finland (KRP) was consulted to ask about the boundaries for the research. One major question was if it was allowed to port scan devices found with the Shodan search engine. When interpreting the law, intentions or goals of the actions have a great emphasis. Port scanning or otherwise inspecting a system for the purpose to break in or exploit the system is explicitly illegal. However, for the purposes of security research, similar actions are not automatically considered illegal; the research was done for the good of the national security and information about found devices was kept confidential and disclosed only to authorities. From this standing point, Kajantie felt that the research was possible but would have to be limited to scanning only the specific ports and not for example digging deeper to find information from within the target systems. Also publicly disclosing the found IP addresses or devices would probably have broken privacy laws, exposing vulnerable data about the targets to the public; it was important to give a full disclosure of the results only to authorities.

Finding out the version number of used software or the version of different services, such as HTTP web server, Telnet and SSH, is legal, if the purpose is only for security research and not to exploit the system. For general purposes finding out software versions is probably not the type of information of which the law on data privacy is meant to protect, so getting that data should not violate any laws.

Querying the service gathers only information that the service advertises, without the need to authenticate or break into the system. That does not violate the article 8 of criminal law chapter 38, that forbids gaining information from inside a computer system. If the service allows logging in to the service without authentication, access can be considered legal. Also gaining HTTP-data prior to authentication is allowed, however, misusing data can break privacy laws even if the data would be legally gained.

Kajantie pointed out some important issues concerning KATSE. She felt that for national security it is better to have a system for scanning operating inside the nation, than using foreign services like Shodan. An own installment of a scanner leaves the queried data inside the nation, and unlike for example Shodan, outside service providers cannot track what kind of devices and vulnerabilities are searched with their service. Kajantie notes, however, that scanning in any case might cause attention among the scanned targets, and she points out that successful communication between involved parties is very important. If the owner of the target system knows about the scan, he will not start investigations on the matter or report the activity to the police. It is also important to disclose the working methods of the scanning system as widely as possible, and emphasize that no damage or disturbances to the systems is being done. Naturally when scanning an entire nation, not everybody can be made aware of the system. If a system owner detects scanning towards his system, he can report it to the police as an attempt for breaking into the system, and the police are legally bound to start investigations. Even though further investigations would quickly prove unnecessary, it still causes additional work and trouble for the system owner and for the police. Thus fluent communication is the key to reduce mistrust and unnecessary investigations.

4.4.2 Legality of KATSE

The intention to run KATSE would be for the good of the national cyber-security. Good intentions alone will not automatically make actions legal, but it is a good starting point when considering the purpose of the system versus national laws. The data learned from the system would not be publicly disclosed or taken advantage of, so privacy laws would not be broken. Also systems would not likely be disturbed or system states altered. The nodes would not break into any systems, being obedient to the §8 of criminal law chapter 38. The nodes simply send packets to open ports and record the responses from the device. That is among the normal functions of communications protocols in any networked device.

It is questionable if retrieving port information, headers and software versions is violating the second moment of §8, which forbids getting information from inside a computer system. The scanning node gets only information that the device is willing to send back, but if the device is misconfigured, it might send information from within a private network to queries from outside sources. Information gained this way would be accidental rather than hacked or on purpose. It remains unclear if the scanner would be considered as a special technical device for exporting information from inside a target system. A possible interpretation of the law is that it is not the

purpose of KATSE to deliberately seek information from inside systems, but just to found out the identity of the device and possible vulnerabilities it has, for national security purposes. Overall, KATSE seems to stand on fairly solid legal grounds, but before using a system like KATSE in public networks, further legal inspection should be made.

Intrusive scanning

If KATSE would do intrusive scanning on systems, the current legislation would need to be changed. Intrusive scanning methods try to break into a system. Even if intrusive scanning would be limited only to bypassing weak security measures such as default credentials, unpatched software or weak encryption, it would still break the current law. If intrusive scanning would be allowed, KATSE should definitely be run by authorities, and security information gained with intrusive methods should remain in their authoritative hands.

To enforce privacy laws, intrusive scans should be as anonymous as possible, only to found out the importance of the target device and its vulnerabilities, and not to learn any other private data from within the target system. If private, identifiable data about individuals would be accidentally gained with intrusive scanning methods, it should not be used for any purposes and should not be recorded or disclosed anywhere. Intrusive scanning methods should be chosen so that no loss of data or system integrity is expected, and the methods should be as harmless as possible. Privacy law is already binding the misuse of confidential information, whether it is gained because of position, by accident or by authoritative rights. For intrusive scanning to be viable, criminal law chapter 38 §8 would have to be modified to give Finnish communications authority (FICORA), or other trusted authority, additional rights to test critical national infrastructure devices for weak security measures such as default credentials and unpatched software exploits.

4.5 Summary

Improving the national cyber-security by finding vulnerable ICS devices automatically is a viable concept. Scanning the entire nation for specific services and identifying the target devices with a fingerprint database is possible: traffic volumes, time consumption of scanning, implementing the scanning software, or the legality of such actions, are not problematic. Some shortcomings of the concept are realized: false negative identifications possibly leave vulnerable devices undetected, identifying the importance of found devices is difficult, and the constant scanning might annoy the target system managers or disturb some sensitive low-end devices. Maintaining a comprehensive ICS device fingerprint database is a crucial part of the concept. Especially for future development, the assessment process for determining the importance and location of found ICS devices, is an important research challenge. Without the assessment, KATSE will produce a lot of results about ICS devices whose role and meaning are vague, requiring human assessment on whether the device is important to national infrastructure or not. On the next chapter, a proof of concept prototype

is presented to validate the identification of devices, traffic volumes, and the speed of scanning when targeting real systems.

5 Proof of concept

In this chapter a proof of concept prototype is presented. The goal of the prototype was to implement some key features of KATSE and try the features against real devices in the Internet. At first, the prototype was tested against a virtual target system, a honeypot, in a private network. The installed honeypot was mimicking a real PLC device of an ICS system with multiple open services, providing a very real-like testing environment. The honeypot is developed by DigitalBond [47] and is designed to be used in the Internet to lure and observe attackers or to distract the attackers away from real systems. After testing with the virtual device, the prototype was given 2913 real IP addresses located in Finland to validate its functions in a real environment and to gather empirical data on found ICS devices, open ports, traffic volume and execution times.

5.1 Prototype

The prototype of KATSE includes most important functions of the concept, as explained below. Programming is done with Python language using also an Nmap library to do the first stage of scanning with Nmap scanning software.

1. Scanning for online hosts and open ports in the assigned IP address list.
2. Querying services based on open ports. At this stage HTTP and Telnet services were implemented.
3. Comparing data from scanning against existing fingerprints for ICS devices.
4. Devices matching ICS device fingerprints are recorded to a text file, containing the device name, IP address, timestamp and the data which led to identification.

The first stage of scanning, i.e. host and port discovery, is done with Nmap using the following options:

```
nmap -sU -sS -p U:137,161,T:21,23,80,443,502,771,8080,3389
```

Host discovery is for determining if the target host is online by sending probes to different ports with different protocols. By default, Nmap uses four methods to do the discovery, which are also the methods used by the prototype: ICMP echo request, a TCP SYN-packet to port 443, a TCP ACK-packet to port 80, and an ICMP timestamp request. The variety of methods is for trying to avoid possible firewalls. Firewalls forwards or drops incoming packets based on pre-set rules; using different methods for scanning increases the odds to get past the firewall and to learn the state of the target host residing behind the firewall. ICMP pings are usually blocked by firewalls; TCP-packet discoveries on the other hand, have a better chance of getting a response from the target. TCP SYN-packet discovery means sending an empty TCP-packet to the target with the SYN-flag set. The SYN-flag suggest to the target host that the other party wishes to establish a TCP-connection. If the target

host is online, it should respond with a SYN/ACK TCP-packet when the target port is open or with a RST-packet when the target port is closed. Either response tells Nmap that the target system is online. The arriving SYN/ACK TCP-packet is dropped by the originating system running Nmap, and a RST-packet is sent to the target to decline further communication. If Nmap does not receive any response, it determines that the target host is offline or unreachable. The same principle as with the SYN-discovery applies to a TCP ACK-packet discovery; an empty TCP-packet is sent to the target with the ACK-flag set, and the target should respond with a RST-packet to terminate the communications resulting from the unexpected ACK-packet. A response from the target reveals to Nmap that the target is online.

The first two options applied to the command line (-sU and -sS) are port scanning, or port discovery, methods, where U stands for User Datagram Protocol (UDP) and S for SYN-packet method with Transmission Control Protocol. The SYN-method is the same method as is used with host discovery. UDP-packets are used to get information from services which usually are used with UDP, mainly SNMP and netBIOS. The SYN-discovery is used because it is fast, does not consume a lot of resources and disturbs the target host as little as possible. Because the SYN-scan drops the connection before the three-level handshake of TCP-connections is complete, it minimizes traffic and might not even show up on the server or service logs on the target host.

The number of ports scanned with Nmap is larger than the amount of services actually implemented in the prototype. For example SNMP, Modbus, RealPort and netBIOS services can be great for identifying ICS devices but they should be used with a certain concern; those services can be very sensitive to outside connections and the methods to communicate with them should be carefully tested before trying against real systems. To make the prototype as safe as possible for the systems it interacts with, the former mentioned services were not implemented, even though previous research [4] indicated that a large number of ICS devices can be identified for example with the SNMP-protocol.

The prototype was able to detect four different ICS devices or product lines according to HTTP- and Telnet-header fingerprints. The three devices or product lines are well-known and popular devices in ICS environments. Table 6 presents the detectable devices; the first and the second entry relate to a specific device, and the third entry to Modicon-line of products by Schneider Automation, which includes large-scale PLCs. The fourth is a heavy-duty industrial serial-to-ethernet network adapter by Westermo.

Table 6: Detectable devices and available fingerprints

Device	Fingerprints
Simatic S7 CP	HTTP
Simatic HMI	Telnet
Schneider Modicon	HTTP, Telnet
Westermo EDW-100/120	HTTP, Telnet

Identification of targets is based on the device fingerprints for different services. In the prototype, the ICS device fingerprint database is formed with text files: referenced device fingerprints are stored in text files, which are read by the prototype and compared to the actual scanning results. Fingerprints are gathered from the previous research [4] as covered earlier in this thesis. Scanning for detailed information from the targets is done by connecting to the Telnet- and HTTP-services if either is found to be open. The prototype tries to connect to the services, records the HTTP- or Telnet-header information and then closes the connection. The process is similar to connecting to a web-site using any web browser. This step is expected to be the most time consuming part of the whole prototype as it is dependent on the responses from the target devices. Especially HTTP-service on a low end target system can be quite slow. Response times from the virtual PLC in the test environment varied from 5 seconds to over 40 seconds, even though the virtual environment was expected to perform steadily. It is unclear why the variation was so tremendous, but a lesson was learned: timeout settings should not be too harsh, giving slow systems enough time to respond.

If the identification process is successful, result of the scan is stored to a text file named after the found device. IP address, timestamp of the scan, and the fingerprint leading to identification, are stored. Additionally, if the found device has an open web-server responding to port 80, an HTML-payload of the hosted web-page is queried and stored. Table 7 presents an example of a result file after an ICS device has been positively identified. The useful piece of information in the example is the "Server"-field which led to identification; HTTP-header reveals the software Decorum which points to Schneider Modicon line of products.

Table 7: Example of recorded information from an identified ICS device

Device name: Schneider_Modicon
Fingerprint match: DECORUM
IP: 123.123.123.123
Timestamp: 1382620446.3414595 (24.10.2013 13:14:6 (UTC))
Scan info:
HTTP/1.0 200 Okay
Date: Sat, 03 Jan 1970 15:36:35 GMT
Connection: close
Content-Length: 4044
Server: DECORUM/2.0
Content-Type: text/html

5.2 Test environment

Honeynets are network segments pretending to be attractive targets for attackers but are actually just luring in attackers to a controlled environment. The SCADA-honeynet by DigitalBond was originally designed to catch intruders and study at-

tacks on ICSs and possibly to protect real ICS assets with distraction. The honeynet includes two virtual machines: one is to simulate a popular Modicon Quantum PLC device from Schneider for luring in attackers, and the other is for observing and managing the honeynet. The observation machine consists of useful tools such as Snort intrusion detection system (IDS) with DigitalBond's own SCADA IDS signatures to monitor network activity. Even though the prototype is only for finding ICS devices, not attacking them, the honeynet provides a great real-like environment for testing. The honeynet PLC is insecure on purpose and has multiple open services willing to respond to queries, which is unfortunately also the case in many real-world industrial systems. For testing the prototype, only the virtual PLC was installed in the test environment, as monitoring or controlling the honeynet was not necessary.

Below Figure 13 illustrates the honeynet. The red box indicating the attacker is also the route where the prototype does the scanning; from the Internet towards a private network. The device in the blue box is a virtual PLC mimicking a real device, and that is the target for scanning. The physical server in the gray box is a laptop running a virtual Ubuntu system. The virtual Ubuntu is the host operating system for the virtual PLC device.

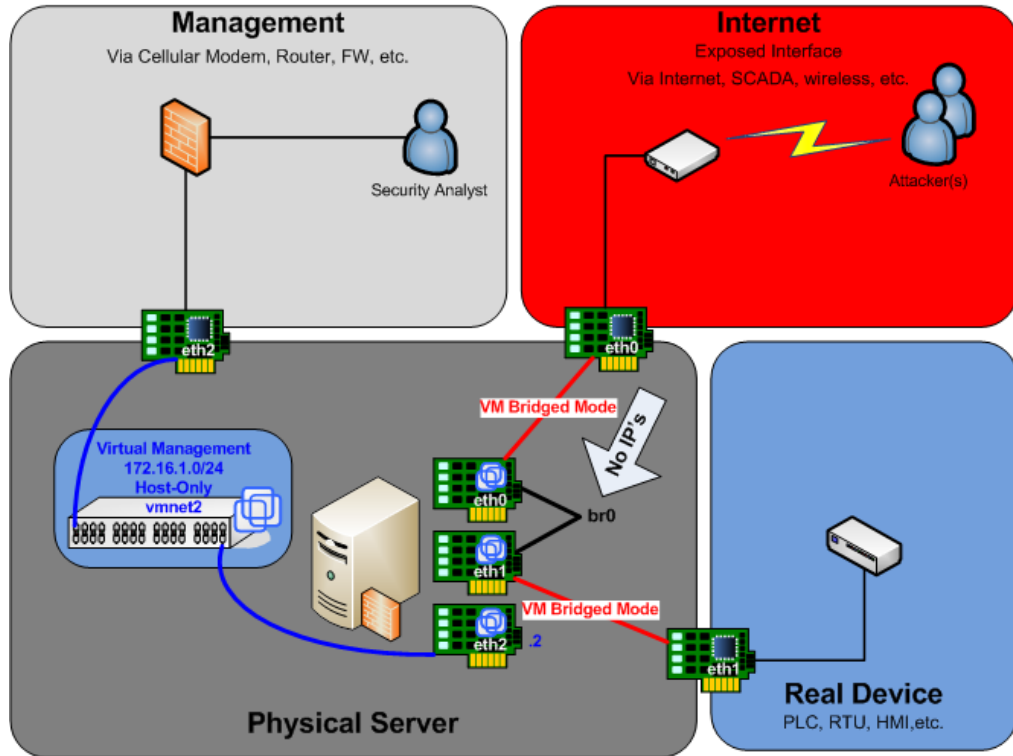


Figure 13: Illustration of the DigitalBond's honeynet environment [47]

5.3 Testing

Before scanning real systems on the Internet, the prototype was tested against a private web-server and a privately hosted virtual PLC as explained earlier. The

test was successful: the first stage of scanning found the targets to be online and correctly determined open ports on the targets. Second stage was able to capture the HTTP-headers from the target hosts and compare the headers to the fingerprints. The prototype correctly identified the virtual PLC as a Schneider Modicon device while not recognizing the regular web-server as an ICS device. Figure 14 presents a caption of the script output from the performed scan.

```

Time of initiation: 28.10.2013 9:33:24 (UTC)
Number of assigned IP-addresses: 2

Starting stage 1 of scanning...
Online hosts found: 2

Starting stage 2 of scanning...
Checking IP: 84.249.██████████...
Schneider_Modicon found from HTTP-fingerprint

Checking IP: 195.148.124.191...
Telnet: timed out
no match found

***Overview***

Hosts and open ports:
{'84.249.██████████': '[161, 21, 23, 502, 8080]', '195.148.124.191': '[23, 80]'}
Number of identified ICS devices: 1

Total time elapsed: 0 hours 0 minutes 15.52 seconds
Script stopped at: 28.10.2013 9:33:40 (UTC)

```

Figure 14: Scan performed against two privately owned hosts

In addition to the overview of the scan the prototype produces a text file showing the status and open ports for every assigned IP address, a text file summarizing online hosts and open ports found, and a result file for every identified ICS device. The result file contains the same info as in the sample in Table 7. Different devices identified with the same fingerprint are distinguishable by the IP address and the timestamp of the scan. A payload-file contains the front page of the hosted web-site in the target host.

A test scan against public web-servers was also made to fish out possible errors in the execution of the prototype. Total of 7 IP addresses were specified: one virtual PLC, one privately hosted web-server, 4 public web-servers such as www.google.fi, and one bogus IP address to have one surely offline host. After sorting out some bugs the prototype worked correctly in this scenario. At this point the prototype was expected to run smoothly against a large number of hosts as well.

5.3.1 Scanning suspected industrial systems in Finland

To validate the prototype in a larger scope and also to gain valuable data on time consumption and traffic volumes, the prototype was run against a selected number of real systems. Because the scanning might be noticed by the administrators of the target systems, a static web-page was hosted at the scanning machine explaining our non-harmful research methods and providing contact information for possible

inquiries. This was done to provide an easy source of information, and to be as transparent as possible with the research.

2913 IP addresses picked from the previous research with Shodan [4] were used as target systems. The prototype was used to scan the same IP addresses on seven different days for observing the consistency of the scans. Number of responding hosts, number of open ports, traffic volume, and time to complete the scanning, and number of identified ICS devices were recorded from each day. The results were somewhat consistent although some variation was observed in the execution time of the scan, in the amount of ICS devices found, and in the amount of responsive hosts in general.

Found ICS devices are categorized in Figure 15, and Table 8 shows the variation of found devices between different dates. The variation between dates suggests that either the hosts are inconsistently reachable, or the prototype is not always able to identify them. As most industrial systems are designed to be constantly on production, it is unlikely that relevant devices would be sometimes offline. A more plausible explanation would be the device's ability to serve only limited number of users at any given time. If all user slots are taken when the prototype tries to query for example the Telnet-service, the target device might automatically terminate the connection due to too many users. One reason for failing identification might be the inability to get the protocol data from the target; a slowly responding host might not always be reached due to low timeout settings in the protocol queries. Used timeout margins in the prototype were 15 seconds for grabbing HTTP-headers and 2 seconds for Telnet-headers.

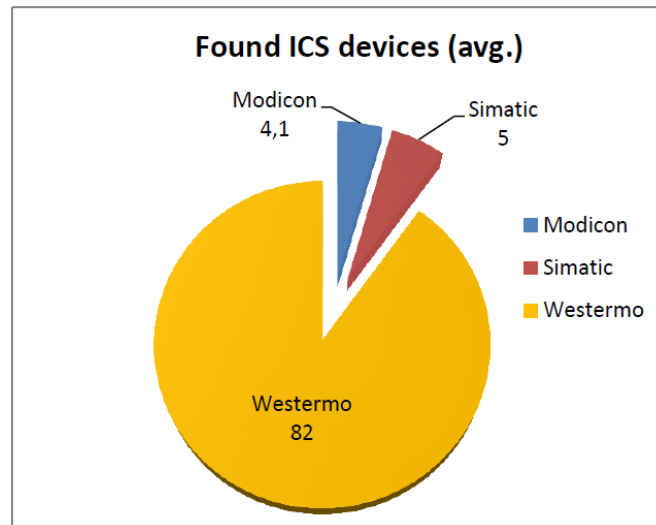


Figure 15: Found ICS devices by brand

The initial Nmap-scan of the prototype to determine online hosts presented also some variation, as can be seen from Figure 16. On average 1845 hosts (63,3%) were found online with a standard deviation of 25,1 hosts between separate scans. The margin between the lowest and highest number of online hosts was 69 (2,4%) which is surprisingly large. Similarly as with the ICS device identification, timeouts are

Table 8: Found devices by date

	12.11.	13.11.	15.11.	18.11.	20.11.	21.11.	22.11.	avg.
Modicon	4	6	2	4	3	4	6	4,1
Simatic	6	6	5	4	5	4	5	5
Westermo	82	80	81	84	84	83	80	82

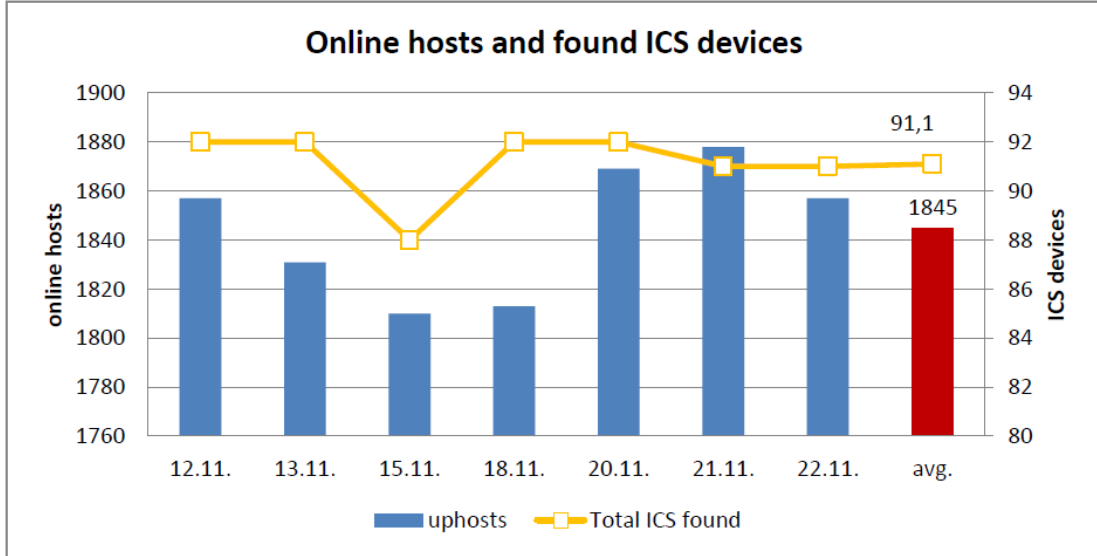


Figure 16: Online hosts and the amount of ICS devices found per day

one possible reason for the variation. However, Nmap is quite persistent and waits patiently for the devices to respond. A more likely explanation is that some hosts are not constantly online, or that their IP address is allocated dynamically which could mean that the previously scanned IP address is not currently allocated to any device or that the IP address is allocated to a different device than before. The yellow curve and the right hand side y-axis in the figure present the total amount of found ICS devices in each scanning attempt.

5.3.2 Traffic and time

Time consumption of the prototype was much higher than expected. Part of the slowness is due to the sequential execution of the prototype: Nmap completes the host discovery for every IP address first, and only after that the second stage of scanning is initiated. By unifying the stages, and doing the scanning with multiple parallel processes, performance would be greatly increased.

Scanning the same IP address list of 2913 hosts took from 3000 seconds to 3735 seconds: an average from seven runs was 3451 seconds meaning the speed of only 0,84 IP addresses per second. That is a surprisingly slow speed compared to the test performed in Chapter 4, where a speed of 45 IP addresses per second was achieved in a local area network. Nmap scans local networks with an ARP-ping discovery method which according to documentation is almost every time faster than the

default methods for non-local networks. Also only 6 hosts were online from the 256-address local network, which is a very low amount of hosts compared to the scanning scheme used against real systems, of which on average 63,3% were online. As the target IP addresses belonged to suspected ICS devices, many of them might be devices with low performance capability and thus answer slower to scan probes than regular computers. Looking closer at the execution time, the first stage of the scanning meaning host and port discovery with Nmap, took approximately 75% of the time, on average 2767 seconds. The Nmap was expected to be much faster than what actually realized, and the second stage of scanning was expected to consume most of the execution time with time consuming HTTP- and Telnet-requests. The second stage took on average 684 seconds meaning 25% of the total time. Also while the execution time of the second stage remained very steady, Nmap was responsible for the big time variations between scans. Nmap's resiliency in waiting for slowly responding hosts and re-sending lost packets might be the issue but in hindsight the slowness of Nmap is not surprising as the execution time was not optimized at all. Even though Nmap scans hosts in parallel, the evidence of greatly varying execution time suggests that the parallelization probably was not working as effectively as it should have. Optimizing the options of Nmap and having a more strict waiting policy, it would probably perform faster but it would also most likely find less online hosts and open ports.

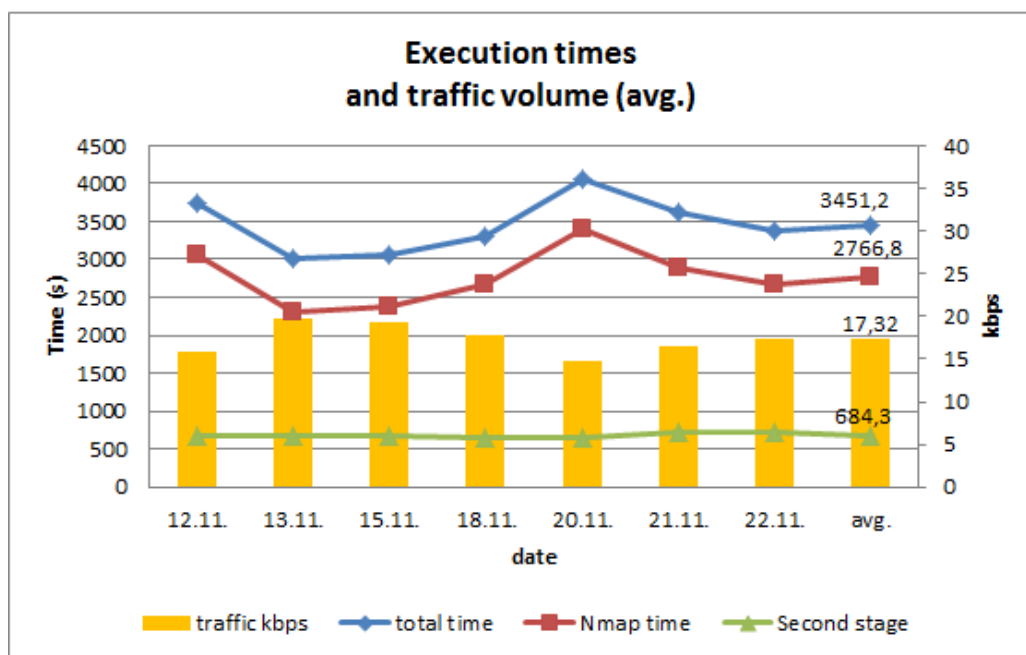


Figure 17: Execution times and the average traffic volume (1 hour = 3600 seconds)

The traffic volumes remained fairly similar between dates. On average, 7,4 MB of total traffic with a standard deviation of 0,05 MB was exchanged between the scanner and the target hosts. The traffic was captured using Tshark packet capture software, filtering out Address Resolution Protocol (ARP) packets as they were not part of the scanning process but produced a noticeable effect on the observed total

Table 9: A summary of scan statistics

	time(s)	traffic(MB)	traffic(kbps)	hosts
min.	3003	7,35	14,82	1810
max.	4049	7,50	19,69	1879
avg.	3451	7,40	17,30	1845
std. dev.	349,5	0,05	1,67	25,1

traffic volume. Figure 17 summarizes the observed traffic volume and execution time of scanning during seven days of testing. An interesting observation is the variation of execution time of the scanner even while the traffic volume remained rather constant. This fact further supports the explanation that apparently Nmap is idling while waiting for slow hosts to respond. Table 9 displays a summary of statistics from the scanning efforts done with the prototype.

5.3.3 Analyzing port numbers

An open port suggests that a service of some kind is listening on that port and accepting connections. Scanning for open ports does not reveal the service running behind the port but as services are usually run behind specific default ports, an educated guess can be made based on the port number. While an open port does not mean that the target device or service is vulnerable, it gives information about possible services running on the device. Analyzing ports can reveal that services used in industrial automation environment are used in the target device, which can help with the importance assessment process of the concept.

From each day of scanning, the amount of open ports found from online hosts were recorded. All separate scans showed similar distribution of open ports so comparing the variation which ports were open seemed unnecessary. To get a wider view of open ports, a separate scan with more ports included was done for the same hosts. In Figure 18 amount of open ports are categorized by port numbers. 'None' suggests that no open ports matching the scanning scheme were found. 543 hosts did not have any ports open, usually indicating that a firewall is probably dismissing the port scanning probes sent by Nmap. Closed ports are actually a good sign; at some point Shodan was able to get information from every host included in the scanning but now only 45% of the targets have open ports. The default services which run on each port were presented earlier in Section 4.2.

If the hosts having zero open ports are not counted, almost 95% of the remaining hosts had the port 80 open suggesting a responsive HTTP-service. Probably quite many of the devices have a web-management interface for remote access, explaining a high percentage of port 80. Ports 502, 771, 135 and 4840 are specifically related to industrial systems. Port 502 is the default port used by a fieldbus protocol called Modbus, which is a quite popular protocol in industrial systems, and a TCP-capable implementation of the protocol exists. Port 771 is the default for RealPort protocol used in serial-to-ethernet servers and gateways, as was discussed in Chapter 3

(Section 3.2.1.). Port 135 is the default for DCOM OPC-protocol which is used to exchange information between automation hardware and hosts using Windows operating system, and is known to have security and configuration related issues. OPC is currently fading from use as OPC/UA-protocol (default port 4840) is replacing it with a more secure architecture including encryption. Having ports used by industrial protocols open, suggests a configuration error, or that the protocols are used over the Internet - both of which can put the target device at risk.

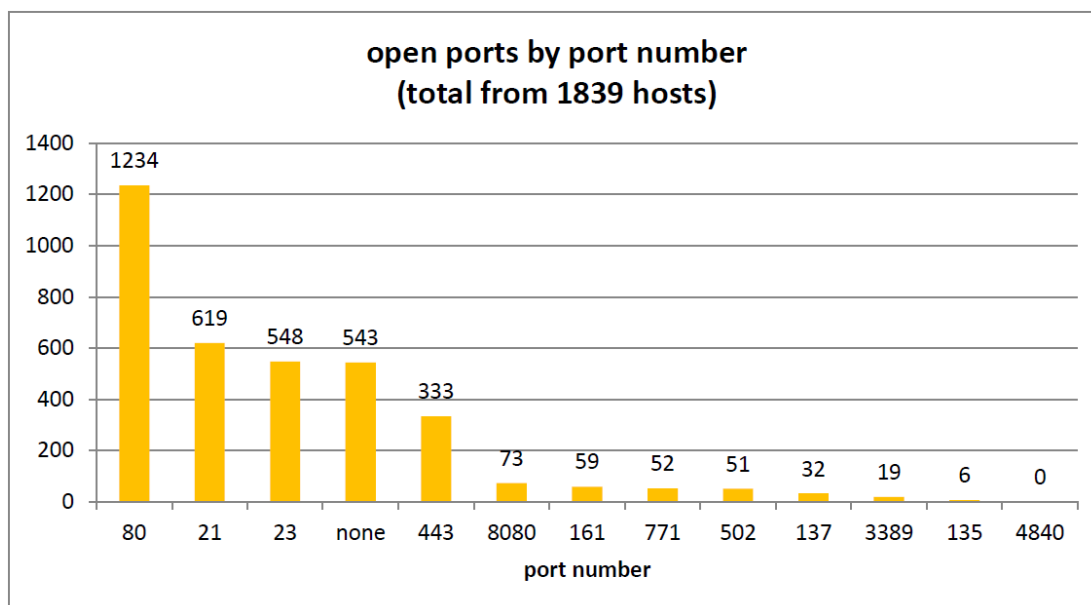


Figure 18: Amount of open ports found from 1839 online hosts

5.4 Traffic and time considerations on a national scale

In Chapter 4, the time to scan every IP address in Finland was calculated from the estimated speed of three different scanners. The estimation for Nmap was 45 hours, which was much faster than what was realized with the prototype. The bright side of the slow performance of the prototype was, that it produced a very small amount of traffic, on average a traffic flow of 17,3 kbps. Increasing scanning speed would also increase the amount of traffic, no matter what software and methods would be used.

Scanning entire nations with millions of IP addresses with the prototype would take months or even years, unless multiple scanner instances would be used in parallel. Table 10 shows how many instances of the prototype would be needed to scan different countries once per day, and once per hour. The amount of IP addresses is based on MaxMind's data [43], and the rest of the table is calculated according to results from this chapter: one prototype instance producing a traffic flow of 17,3 kbps and scanning 0,84 IP addresses per second. The number of instances needed is quite substantial. Nevertheless, each instance consumes very little memory and processing power, enabling the use of hundreds of instances per one powerful server.

Table 10: Scanning every IP address in a nation once per day, and once per hour with multiple instances of the scanner.

IP addresses	inst. (1/d)	traf. (1/d)	inst. (1/h)	traf. (1/h)
Finland, 13,6M	188	3,3 Mbps	4512	79,2 Mbps
Sweden, 27,1M	374	28,8	8976	156,0
Russia, 46,4M	639	22,9	15336	266,4
France, 96,1M	1324	11,1	31776	549,6
Germany, 120,9M	1665	6,5	39960	691,2
USA, 1580,6M	21778	376,8	522672	9043,2

During execution, the memory consumption of the prototype was around 35 MB: a well-equipped server with 32 GB of RAM could host 914 of such instances. Traffic-wise having 914 instances in one location would not be a problem: only 15,8 Mbps of traffic would be induced. Looking at the table, Finland, Sweden and Russia, could be scanned with a single server in the case of scanning every address once per day. Scanning France, Germany, and the USA, along with other countries in the case of once per hour scanning would require multiple physical servers located in different networks. The USA stands out from the rest of the countries with the need of 522672 scanner instances in order to scan the nation once per hour; a result from the USA having a huge number of allocated IP addresses. 572 servers, each hosting 914 scanner instances, would be required to achieve that goal for the USA.

As can be seen from the table, traffic volumes are quite moderate even with thousands of instances of the prototype; scanning Finland in an hour with 4512 instances would only produce a constant traffic flow of 79,2 Mbps, constituting for only 0,5% of the average Internet traffic in Finland. USA, having a monstrous amount of allocated IP addresses, is the only nation which could not be scanned once per hour with a high-speed 1 gigabit per second network link; required 9 Gbps of traffic should be distributed to several networks to avoid congestion. Like in Finland, the traffic volume by the prototype instances would not be a burden for the network infrastructure. For example, the Internet exchange point DE-CIX in Frankfurt, Germany, gets on average 1535,4 Gbps of network traffic [48] (yearly average). The traffic flow of 691,2 Mbps needed to scan Germany would be really insignificant in comparison.

Calculations in this section suggest that Nmap is probably not the ideal software for scanning large chunks of IP addresses quickly. But it is important to note that while the ZMap scanner is advertised to scan the entire Internet (3,6 billion addresses) in 45 minutes [35], that only accounts for a single port in each host; the prototype, utilizing Nmap, scans for 10 ports, and additionally queries time consuming HTTP- and Telnet-request from the hosts. Even if host and port discoveries would be done very quickly with ZMap, querying service information, which is essential for identifying the hosts, would be the slowest part of the system. The prototype managed to do Telnet- and HTTP-queries with the average speed of 4,26 IP addresses per second. With that speed, roughly 37 instances would still be needed

to scan every IP address in Finland once a day.

5.5 Summary

The prototype successfully used data from port scanning and pre-recorded fingerprints to identify ICS devices of four different kinds from a given IP address pool. The prototype was tested against a virtual ICS device before scanning real systems, and information was only gathered with HTTP- and Telnet-protocols as they are very safe to use; protocols such as SNMP and Modbus are valuable for finding devices but they may also be harmful and disruptive to the device being scanned.

Scanning real systems provided interesting data on found ICS devices, open ports, traffic volume and execution time. IP address list which was scanned consisted of 2913 IP addresses and on average 91 of them were identified as ICS devices. Traffic volume was confirmed to be unnoticeable in a national perspective; the prototype produced on average 17 kbps constant data stream, while the Finnish Internet traffic is around 18 Gbps. The scanning speed of the prototype was very slow, 0,84 hosts per second, but having multiple parallel scanners or using a faster port scanning software would provide a huge increase in speed making the daily or weekly scanning of every IP address in Finland possible. Even scanning much larger nations than Finland once per day, or once per hour, would not be a burden on the network infrastructure of the nations'. Running hundreds of instances of the prototype in a well-equipped server, enables a high scanning speed with moderate traffic volume. However, currently the memory consumption of the instances is the bottle-neck when utilizing a single server.

An interesting observation was the variation on the amount of online hosts each day. The prototype was used on seven different days to scan the same IP addresses and the count for online hosts varied from 1810 to 1879. Possible reasons for this are slowly responding target hosts, timeout settings in the prototype, dynamically changing IP addresses or devices which are not constantly turned on. More research on the mentioned explanations would be needed to make any conclusions but as the target IP addresses belong to suspected ICS devices, it is unlikely that they would be inconsistently available.

6 Conclusions

Nations worldwide have realized the importance of cyber-security in a modern society. Highly networked infrastructure elements such as power grids and waste water treatment facilities are reachable from the Internet and need to be appropriately protected. The Finnish cyber-strategy [1] states that cooperation between relevant actors in the cyber-space is the key to get information about the state of the nation's cyber-space; for example Cert-fi, Internet service providers, companies part of the supply of critical infrastructure, researchers, and private security companies should share information to achieve a timely situational awareness of possible threats the nation is facing.

Industrial control systems have traditionally been isolated systems with the emphasis on functionality and safety of personnel and equipment; there was no need for information security. However in the last decade an increasing trend of unifying system architectures and enabling remote management for economic reasons have exposed ICSs to environments where their security is not on par with the current challenges of information security in the Internet. Systems are exposed to Internet without proper protection which is a highly worrying situation; the potential of ICS failures can be even catastrophic as was seen with the explosion of Tsernobyl [19] nuclear power plant or a dam explosion in Russia [18] resulting in multiple deaths.

Improving productivity of ICSs is understandable but security policies should be updated to modern standards. Enumerating ICS devices from the Internet is possible and rather easy; systems cannot rely on security by obscurity anymore. Several studies have implicated that real industrial control systems are exposed to Internet and attacks towards them are happening [20, 31, 4]. A security researcher in the U.S. observed highly skilled targeted attacks towards the fake water power plant he had setup with honeypot devices; attacks originated from many countries and some of the attackers were savvy enough to take a complete control over the system [21].

KATSE and the proof of concept

On a national level the communications security authority could do port scanning inside the nation to find and remediate vulnerable ICSs. In this thesis a concept is proposed to find vulnerable ICS devices automatically, analyzing the importance of the devices and reporting them to authorities for securing the most important and critical devices quickly. The concept, called KATSE, would scan every Internet connected device inside Finland once a day to find exposed ICSs. KATSE works without disturbing the scanned device, and on a long term, it could be used to get rid of accidentally exposed industrial devices, effectively decreasing the attack surfaces towards the critical infrastructure and consequently improving the nation's cyber-security. Additionally, proving that ICS devices can be found might raise security awareness among the industry and result in better security practices.

Challenges with the concept relate to analyzing the importance of found ICS devices, legal aspects preventing thorough scanning, and a practical matter of reaching

the owners of the found devices and actually securing the devices. The owners of the devices might not acknowledge the threats or just simply will not act on the matter. From the 1968 online ICS and building automation devices found in the Shodan research in quarter one of 2013 [4], 1602 was still found to be reachable roughly 7 months later even though a report of the original findings were given to the authorities. This is evidence to the fact that getting information through to the owners of found devices and improving the security of those devices is indeed a challenge.

A proof of concept prototype was constructed to validate the key functions of KATSE. The prototype was programmed to be able to find four different kinds of ICS devices based on pre-recorded fingerprint data. The same 2913 IP addresses found in the Shodan-research were scanned with the prototype: from seven runs, on average 91 ICS devices were found from an average of 1845 online hosts. Data about traffic volume, execution time and open ports were also recorded, validating that traffic volume produced by the scanner is unnoticeable compared to overall Finnish Internet traffic. Even though the prototype was quite slow, the speed of scanning can be easily improved; the daily scanning of an entire nation of the size of Finland is easily possible.

Future work and challenges

Key improvements for KATSE would be the ability to accurately analyze the importance of found ICS devices and to introduce a thorough security auditing of the found devices. From thousands of devices which might be found, determining critical devices automatically would save considerable amount of time and effort from personnel and it would allow a quick remediation process for devices part of critical infrastructure. An automated security auditing for the devices could also help in assessing the importance and to get an estimate of how vulnerable the found device is. However, changes in legislation would be needed as currently any kind of unauthorized security auditing such as trying out default credentials, is forbidden. In the future, changes in legislation could be introduced to allow authoritative entities to perform security auditing for critical infrastructure systems for the sake of national security, keeping in mind the possible sensitivity of the target systems, and various privacy concerns.

If a third-party security auditing for critical devices would be legalized for authorities, expanding the reach of KATSE could prove useful; for example finding security flaws from building automation systems in hospitals or inspecting backbone routers and other important systems could help in finding weak spots in the nation's cyber-space.

References

- [1] *Suomen kyberturvallisuusstrategia: Valtioneuvoston periaatepäätös 24.1.2013.* Puolustusministeriö, 2013.
- [2] J. Weiss, *Protecting industrial control systems from electronic threats.* Momentum Press, 2010.
- [3] N. C. System, “Supervisory Control and Data Acquisition (SCADA) systems.” Online: http://www.ncs.gov/library/tech_bulletins/2004/tib_04-1.pdf, October 2004. Retrieved Aug 8, 2013.
- [4] S. Tiilikainen and J. Manner, “Suomen automaatioverkkojen haavoittuvuus - raportti Internetissä julkisesti esillä olevista automaatiolaitteista.” Online: <https://research.comnet.aalto.fi/public/Aalto-Shodan-Raportti-julkinen.pdf>, March 2013. Retrieved Aug 20, 2013.
- [5] J. N. Hoover, “Thousands of industrial control systems at risk: DHS study.” Online: <http://www.informationweek.com/government/security/thousands-of-industrial-control-systems/240146091>, January 2013. Retrieved Aug 19, 2013.
- [6] K. S. Keith Stouffer, Joe Falco, “Guide to industrial control systems (ICS) security,” *Recommendations of the National Institute of Standards and Technology, Special Publication 800-82*, 2011.
- [7] K. S. Keith Stouffer, Joe Falco, “Recommended practise: Improving industrial control systems cybersecurity with defense-in-depth strategies,” *Department of Homeland Security, Control systems security program, national cyber security division*, 2009.
- [8] E. Byres and J. Lowe, “The myths and facts behind cyber security risks for industrial control systems,” in *Proceedings of the VDE Kongress*, vol. 116, 2004.
- [9] ICS-CERT, “Medical devices hard-coded passwords.” Online: <http://ics-cert.us-cert.gov/alerts/ICS-ALERT-13-164-01>, June 2013. Retrieved Aug 20, 2013.
- [10] DigitalBond, “Schneider Modicon Quantum.” Online: <http://www.digitalbond.com/tools/basecamp/schneider-modicon-quantum>. Retrieved Aug 20, 2013.
- [11] National Institute of Standards and Technology, “Guide to industrial control systems (ICS) security,” 2011.
- [12] R. Johnson, “Survey of SCADA security challenges and potential attack vectors,” in *Internet Technology and Secured Transactions (ICITST), 2010 International Conference for*, pp. 1 –5, nov. 2010.

- [13] V. Iguire, S. Laughter, and R. Williams, "Security issues in SCADA networks," *Computers & Security*, vol. 25, no. 7, pp. 498–506, 2006.
- [14] É. P. Leverett, "Quantitatively assessing and visualising industrial system attack surfaces," *Master's thesis, University of Cambridge*, 2011.
- [15] J. Matherly, "ShodanHQ." Online: <http://www.shodanhq.com>. Retrieved Aug 19, 2013.
- [16] ERIPP, "Every routable IP project." Online: <http://www.eripp.com>. Retrieved Sep 2, 2013.
- [17] S. strangelove. Online: <http://scadastrangelove.org>. Retrieved Oct 11, 2013.
- [18] J. P. Hasler, "Investigating Russia's biggest dam explosion: what went wrong." Online: <http://www.popularmechanics.com/technology/engineering/gonzo/4344681>, February 2010. Retrieved Jul 16, 2013.
- [19] A. Nicholson, S. Webber, S. Dyer, T. Patel, and H. Janicke, "SCADA security in the light of cyber-warfare," *Computers & Security*, vol. 31, no. 4, pp. 418 – 436, 2012.
- [20] E. J. Markey, "Electric grid vulnerability - industry responses reveal security gaps." Online: http://markey.house.gov/sites/markey.house.gov/files/documents/Markey%20protect%20T1\textdollarGrid%20Report_05.21.13.pdf, May 2013. Retrieved Jul 10, 2013.
- [21] T. Simonite, "Chinese hacking team caught taking over decoy water plant." Online: <http://www.technologyreview.com/news/517786/chinese-hacking-team-caught-taking-over-decoy-water-plant>, August 2013. Retrieved Aug 6, 2013.
- [22] M. I. C. Report, "APT1: Exposing one of China's cyber espionage units." Online: http://intelreport.mandiant.com/Mandiant_APT1_Report.pdf, February 2013. Retrieved Aug 6, 2013.
- [23] T. Grenman, "Case: tietojä varastava haittaohjelma." Online: https://www.cert.fi/attachments/cipseminarit/cip_2012/6Bsh0S3jp/Grenman.pdf, 2012. Retrieved Sep 28, 2013.
- [24] P. Bright, "U.S. agency baffled by modern technology, destroys mice to get rid of viruses." Online: <http://arstechnica.com/information-technology/2013/07/us-agency-baffled-by-modern-technology-destroys-mice-to-get-rid-of-viruses>, July 2013. Retrieved Jul 18, 2013.
- [25] R. Hughes, "A treaty for cyberspace," *International Affairs*, vol. 86, pp. 523+, MAR 2010.

- [26] H. Tibbs, "The global cyber game." Online: http://www.da.mod.uk/publications/library/technology/20130508-Cyber_report_final_U.pdf, May 2013. Retrieved Aug 12, 2013.
- [27] U.S. Department of Defense, Washington D.C.: U.S. Joint Chiefs of Staff, "National military strategy for cyberspace operations", 2006.
- [28] M. N. Schmitt, ed., *Tallinn Manual on the International Law Applicable to Cyber Warfare*. Cambridge University Press, 2013.
- [29] S. Gorman, "Electricity grid in U.S. penetrated by spies." Online: <http://online.wsj.com/article/SB123914805204099085.html>, April 2009. Retrieved Jul 18, 2013.
- [30] R. J. Deibert, R. Rohozinski, and M. Crete-Nishihata, "Cyclones in cyberspace: Information shaping and denial in the 2008 Russia–Georgia war," *Security Dialogue*, vol. 43, no. 1, pp. 3–24, 2012.
- [31] D. E. Sanger, "Obama order sped up wave of cyberattacks against Iran," *New York Times*, June 2012.
- [32] N. Falliere, L. Murchu, and E. Chien, "W32.Stuxnet dossier." Online: http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/w32_stuxnet_dossier.pdf, 2010. Retrieved Oct 24, 2011.
- [33] O. Dictionaries, "Definition of Internet of Things in English." Online: <http://oxforddictionaries.com/definition/english/Internet-of-things>. Retrieved Sep 19, 2013.
- [34] R. Shirey, "RFC 2828: Internet Security Glossary," May 2000. Status: Informational.
- [35] Z. Durumeric, E. Wustrow, and J. A. Halderman, "ZMap: Fast Internet-wide scanning and its security applications," in *Proceedings of the 22nd USENIX Security Symposium*, Aug. 2013.
- [36] R. Graham, "Masscan: Mass IP port scanner." Online: <https://github.com/robertdavidgraham/masscan>. Retrieved Sep 19, 2013.
- [37] A. author, "Internet Census 2012." Online: <http://internetcensus2012.bitbucket.org/paper.html>, January 2013. Retrieved Aug 19, 2013.
- [38] H. Moore, "Serial offenders: widespread flaws in serial-port servers." Online: <https://community.rapid7.com/community/metasploit/blog/2013/04/23/serial-offenders-widespread-flaws-in-serial-port-servers>, 2013. Retrieved Aug 16, 2013.

- [39] J. Matherly, "Internet cartography." Online: https://www.swisscyberstorm.com/files/3013/7210/4071/Internet-Cartography_John-Matherly.pdf, June 2013. Retrieved Jul 17, 2013.
- [40] D. Goodin, "Defects leave critical military, industrial infrastructure open to hacks." Online: <http://arstechnica.com/security/2012/07/ics-security-light-years-behind-itunes/>, July 2012. Retrieved Jul 17, 2013.
- [41] J. Kenttälä, "HAVARO." Online: https://www.cert.fi/attachments/cipseminaarit/cip_2012/6Bsh7Xlim/Kenttala.pdf, 2012. Seminar on Critical Infrastructure Protection. Retrieved Jun 28, 2013.
- [42] BBC, "Sasser net worm affects millions." Online: <http://news.bbc.co.uk/1/hi/technology/3682537.stm>, May 2004. Retrieved Jan 13, 2014.
- [43] MaxMind. <http://www.maxmind.com/en/techinfo>, August 2013. Retrieved Aug 27, 2013.
- [44] SANS, "Intrusion detection FAQ: What is Scanrand?." Online: <http://www.sans.org/security-resources/idfaq/scanrand.php>. Retrieved Aug 28, 2013.
- [45] RIPE Network Coordination Center. <https://www.ripe.net/data-tools/db/about-the-ripe-database>, August 2013. Retrieved Aug 27, 2013.
- [46] MaxMind. <http://www.maxmind.com/en/city>, August 2013. Retrieved Aug 27, 2013.
- [47] DigitalBond, "SCADA honeynet." Online: <http://www.digitalbond.com/tools/scada-honeynet/>, 2006. Retrieved Oct 23, 2013.
- [48] DE-CIX, "Statistics." Online: <http://www.de-cix.net/about/statistics/>. Retrieved Jan 15, 2014.